# Partially observed optimal stopping problem for discrete-time Markov processes

Benoîte de Saporta, François Dufour, Christophe Nivot

# Partially observed optimal stopping problem for discrete-time Markov processes

B. de Saporta
Université de Montpellier, France
IMAG, CNRS UMR 5149, France
INRIA Bordeaux Sud Ouest
e-mail : benoite.de-saporta@umontpellier.fr

F. Dufour *
Bordeaux INP, France
IMB, CNRS UMR 5251, France
INRIA Bordeaux Sud Ouest
e-mail : dufour@math.u-bordeaux1.fr

C. Nivot
INRIA Bordeaux Sud Ouest
Université de Bordeaux, France
IMB, CNRS UMR 5251, France
e-mail : christophe.nivot@inria.fr

February 16, 2016

**Abstract**

This paper is dedicated to the investigation of a new numerical method to approximate the optimal stopping problem for a discrete-time continuous state space Markov chain under partial observations. It is based on a two-step discretization procedure based on optimal quantization. First,we discretize the state space of the unobserved variable by quantizing an underlying reference measure. Then we jointly discretize the resulting approximate filter and the observation process. We obtain a fully computable approximation of the value function with explicit error bounds for its convergence towards the true value fonction.

**Keywords:** Optimal stopping, partial observations, Markov chain, dynamic programming, numerical approximation, error bound, quantization.

**AMS 2010 subject classification:** 60J05, 60G40, 93E11

## 1 Introduction

This paper is dedicated to the investigation of a new numerical method to approximate the optimal stopping problem for a discrete-time continuous state space Markov chain under partial observations. This is known to be a difficult problem, but very important for practical applications. Indeed, the usual approach when dealing with partially observed problems is to introduce the filter or belief process, thus converting the problem into a fully observed one, at the cost of an infinite

*Corresponding author: F. Dufour, INRIA, 200 Avenue de la Vieille Tour, 33405 Talence Cedex, France.

dimensional state space as the filter process is measure-valued. Thus, there is no straightforward way to discretize the state space of the filter process, and one must often choose a balance between the computational load and the accuracy of the approximation.

Unlike the huge literature on discrete state space optimal stopping problems, that on continuous state space is scarce. The most relevant papers addressing this problem are [8, 9, 10]. In [8], the authors do not propose an approximation of the value function, but only computable upper and lower bounds. They do not require any particular assumptions on the Markov process apart from being simulatable, however they do not provide convergence rates either. Our aim in this paper is more ambitious as we want to construct a numerically tractable approximation of the value function with a bound for the convergence rate. In [9], the authors compute an approximation of the value function based on particle filtering and simulations of the chain trajectories. They assume that all distributions have densities with respect to the Lebesgue measure, that the reward function is convex, but again they don't provide convergence rates for the approximation. Our approach is more general as it allows the Markov chain kernel have a density with respect to a general product of measures, not necessarily the Lebesgue measure, which may be more relevant for some applications where thresholds are involved, for instance. In [10], the authors propose to parametrize the belief state with the exponential family to dramatically reduce its dimension, but they deal with general control problems for infinite discounted cost and stationary policies that are not suitable for optimal stopping problems.

In this paper we propose a new approach, inspired by [7] that addresses the optimal stopping problem under partial observation for finite state space chains. The key idea of the authors is to approximate simultaneously the filter and observation processes using a series of quantization grids. Optimal quantization is an approximation procedure that replaces a continuous state space variable $X$ by a finite state space one $\widehat{X}$ optimally, in the sense that it minimizes the $L_2$ norm of the difference $|X - \widehat{X}|$, see e.g. [5, 6] and references therein for more details and applications to numerical probability. The quantization approach of [7] is especially efficient if the state space of the unobserved variable is finite and small. One first simple idea to turn our continuous state space problem into a discrete one is to discretize the state space of the unobserved variables using a regular cartesian grid. However, to ensure precision this may require a huge number of points and possibly useless computations if some areas of the state space are seldom visited. A better idea is to use the same quantization approach as [7] to discretize the unobserved component. This will ensure that the grids have more points in the areas of high density, and are dynamically adapted with time. The state space of the unobserved variable would then be finite, but with time-varying, making the discretization of the filter numerically intractable. Our approach attempts at taking the advantages of both these ideas, while minimizing their drawbacks. Our approximation procedure is in two steps. First, we discretize the state space of the unobserved variable by optimally quantizing an underlying reference distribution. Thus we have a fixed finite state space for the unobserved variable, and the points are optimally distributed to ensure precision at a minimal computational cost. This yields an approximate filter process that is measure-valued, but can be seen as taking values in a finite dimensional simplex. We then jointly quantize the approximate filter and observation processes. Throughout this procedure, we are able to compute an explicit upper bound for the error, that goes to zero as the number of points in the quantization grids goes to infinity.

The paper is organized as follows. In Section 2, we state the optimal stopping problem under partial observation we are intereted in approximating, and we give the equivalent completely observed sequential decision making problem. In Section 3, we detail our two-step numerical scheme

and evaluate the error bound. Section 4 is dedicated to a numerical example, and the most technical results are postponed to an Appendix.

# 2 Problem formulation

We start with some general notation that will be in force throughout the paper. $\mathbb{N}$ is the set of natural numbers including 0, $\mathbb{N}^* = \mathbb{N} - \{0\}$, $\mathbb{R}$ denotes the set of real numbers, $\mathbb{R}_+$ the set of non-negative real numbers, $\mathbb{R}_+^* = \mathbb{R}_+ - \{0\}$. For any $(p,q) \in \mathbb{N}^2$ with $p \leq q$, $[\![p;q]\!]$ is the set $\{p, p+1, \ldots, q\}$. Given $x$ in the Euclidean space $\mathbb{R}^n$,$|x|$ will denote its Euclidean norm. Let $I_E$ be the indicator function of a set $E$. Let $E$ be a metric space where $d$ denotes its associated distance. Its Borel $\sigma$-algebra will be denoted by $\mathcal{B}(E)$ and $\mathcal{P}(E)$ is the set of probability measures on $(E, \mathcal{B}(E))$. The space of all bounded real-valued measurable function on $E$ is denoted by $\mathbb{B}(E)$. The space $\mathbb{L}(E)$ of all real-valued bounded Lipschitz continuous functions on $E$ is equipped with the norm $\|f\|_{\mathbb{L}(E)} = \|f\|_{sup} + L_f$ where $\|f\|_{sup} = \sup_{x\in E} |f(x)|$ and $L_f = \sup_{x\neq y} \frac{|f(x)-f(y)|}{d(x,y)}$ and $\mathbb{L}_1(E) = \{f \in \mathbb{L}(E) : \|f\|_{\mathbb{L}(E)} \leq 1\}$. On $\mathcal{P}(E)$, let us introduce the distance $d_{\mathcal{P}}$ defined by $d_{\mathcal{P}}(\mu,\nu) = \sup_{f\in\mathbb{L}_1(E)} \left\{ \int_E f d\mu - \int_E f d\nu \right\}$. The Dirac probability measure concentrated at $x \in E$ will be denoted by $\delta_x$. If $F$ is a metric space and $v$ is a real-valued bounded measurable function defined on $E \times F$ and $\gamma$ is a probability measure on $(E, \mathcal{B}(E))$ then by a slight abuse of notation we write $v(\gamma, y) = \int_E v(x,y)\gamma(dx)$ for any $y \in F$.

## 2.1 Optimal stopping

In this section, we describe the optimal stopping problem we are interested in by using a *weak* formulation. Consider $\mathbb{X} \in \mathcal{B}(\mathbb{R}^m)$, $\mathbb{Y} \in \mathcal{B}(\mathbb{R}^n)$, a stochastic kernel $R$ on $\mathbb{X} \times \mathbb{Y}$ and a performance function $\mathbf{H} \in \mathbb{B}(\mathbb{X} \times \mathbb{Y})$.

**Definition 2.1** *The control is defined by the following term:*

$$\ell \quad = \quad \big(\Xi, \mathcal{G}, \mathbf{Q}, \{\mathcal{G}_t\}_{t\in[\![0;N_0]\!]}, \{\boldsymbol{\mathcal{X}}_t, \boldsymbol{\mathcal{Y}}_t\}_{t\in[\![0;N_0]\!]}, \tau\big)$$

- $\big(\Xi, \mathcal{G}, \mathbf{Q}, \{\mathcal{G}_t\}_{t\in[\![0;N_0]\!]}\big)$ *is a filtered probability space,*
- $\{\boldsymbol{\mathcal{X}}_t, \boldsymbol{\mathcal{Y}}_t\}_{t\in[\![0;N_0]\!]}$ *is an* $\mathbb{X}\times\mathbb{Y}$*-valued* $\{\mathcal{G}_t\}_{t\in[\![0;N_0]\!]}$*-Markov chain defined on* $\big(\Xi, \mathcal{G}, \mathbf{Q}\big)$ *where* $R$ *is its associated transition kernel and* $\delta_{(\mathbf{x},\mathbf{y})}$ *is its initial distribution,*
- $\tau$ *is a* $\{\mathcal{G}_t^{\boldsymbol{\mathcal{Y}}}\}_{t\in[\![0;N_0]\!]}$*-stopping time where* $\mathcal{G}_t^{\boldsymbol{\mathcal{Y}}} = \sigma\{\boldsymbol{\mathcal{Y}}_0, \ldots, \boldsymbol{\mathcal{Y}}_t\}$.

In this setting, $\boldsymbol{\mathcal{X}}_t$ denotes the hidden variables and $\boldsymbol{\mathcal{Y}}_t$ the observed variables. Hence the stopping decision $\tau$ depends only on the observations. The set of the previous controls is denoted by $L$ and $\boldsymbol{\mathcal{E}}^{\mathbf{Q}}$ denotes the expectation under the probability $\mathbf{Q}$. For a control $\ell \in L$, the performance criterion is given by

$$\boldsymbol{\mathcal{H}}(\mathbf{x}, \mathbf{y}, \ell) \quad = \quad \boldsymbol{\mathcal{E}}^{\mathbf{Q}}\big[\mathbf{H}(\boldsymbol{\mathcal{X}}_\tau, \boldsymbol{\mathcal{Y}}_\tau)\big]. \tag{1}$$

In the previous expression, we write explicitly the dependence of the cost function on the initial state of the Markov chain. The partially observed optimal stopping problem we are interested in is to maximize the reward function $\boldsymbol{\mathcal{H}}(\mathbf{x}, \mathbf{y}, \ell)$ over $L$. The corresponding value function is thus

$$\overline{\boldsymbol{\mathcal{H}}}(\mathbf{x}, \mathbf{y}) = \sup_{\ell\in L} \boldsymbol{\mathcal{H}}(\mathbf{x}, \mathbf{y}, \ell). \tag{2}$$

The aim of this paper is to propose a numerical approximation of $\overline{\boldsymbol{\mathcal{H}}}(\mathbf{x}, \mathbf{y})$ that can be computed in practice and derive bounds for the approximation error.

We make the following main assumptions on the parameters of the Markov chain and the performance function. The first ones are mild and state that the transition kernel $R$ of the Markov chain has a density with respect to a product of reference probability measures, and that this density is bounded with Lipschitz regularity. Assumption C is technical and more restrictive. It states that the density should also be bounded from below. Finally we assume that the performance function is also bounded and Lipschitz-continuous.

**Assumption A.** There exist $\lambda \in \mathcal{P}(\mathbb{X})$, $\nu \in \mathcal{P}(\mathbb{Y})$ and an $\mathbb{R}_+$-valued measurable function $r$ defined on $(\mathbb{X} \times \mathbb{Y})^2$ such that

*(A1) for any* $(x, y) \in \mathbb{X} \times \mathbb{Y}$, $B \in \mathcal{B}(\mathbb{X})$, $C \in \mathcal{B}(\mathbb{Y})$

$$R(B \times C|x, y) = \int_{B \times C} r(u, v, x, y)\lambda(du)\nu(dv),$$

*(A2)* $\int_{\mathbb{X}} |x|^{2+\beta}\lambda(dx) < \infty$ *for some* $\beta > 0$.

**Assumption B.** There exist positive constants $\overline{r}$ and $L_r$ such that

*(B1)* $\sup_{(u,v,x,y)\in(\mathbb{X}\times\mathbb{Y})^2} r(u, v, x, y) \leq \overline{r}$

*(B2) for any* $(u, v, x, y) \in (\mathbb{X} \times \mathbb{Y})^2$, $(u', x', y') \in \mathbb{X} \times \mathbb{X} \times \mathbb{Y}$

$$|r(u, v, x, y) - r(u', v, x', y')| \leq L_r\big[|u - u'| + |x - x'| + |y - y'|\big].$$

**Assumption C.** There exists $\delta > 0$ such that $r(\lambda, v, x, y) \geq \delta^{-1}$ for any $(v, x, y) \in \mathbb{Y} \times \mathbb{X} \times \mathbb{Y}$.

**Assumption D.** The function $\mathbf{H}$ belongs to $\mathbb{L}(\mathbb{X} \times \mathbb{Y})$.

Our approximation strategy is in three steps. First, we rewrite the problem as a sequential decision-making problem for a fully observed Markov chain on $\mathcal{P}(\mathbb{X}) \times \mathbb{Y}$. Then we propose a first approximation based on the discretization of the state space $\mathbb{X}$ by a finite grid $\Gamma_X^N$. Finally, we use a second approximation procedure to discretize the resulting Markov chain on $\mathcal{P}(\Gamma_X^N) \times \mathbb{Y}$.

## 2.2 Auxiliary completely observed control problem

As explained in the introduction, the standard approach to deal with partial observation is to introduce the filter process and convert the problem into a fully observed one on an infinite dimensional state space. In this section, we introduce the auxiliary completely observed control model $\mathcal{M}$. We follow closely the framework of Chapter 5 in [2]. The objective of this section is twofold. First, we show that the optimal stopping problem introduced in Definition 2.1 is equivalent to a fully observed optimization problem defined in terms of the control model $\mathcal{M}$ below. Second, we prove that the value function $\overline{\boldsymbol{\mathcal{H}}}(\mathbf{x}, \mathbf{y})$ defined in (2) can be obtained by iterating a Bellman operator.

As defined in [2], let us consider the Bayes' operator $\Phi : \mathbb{Y} \times \mathcal{P}(\mathbb{X}) \times \mathbb{Y} \mapsto \mathcal{P}(\mathbb{X})$ given by $\Phi(v, \theta, y)(du) = \dfrac{r(u, v, \theta, y)}{r(\lambda, v, \theta, y)}\lambda(du)$ and the stochastic kernel $S$ on $\mathcal{P}(\mathbb{X}) \times \mathbb{Y}$ defined by

$$S(B \times C|\theta, y) = \int_C \delta_{\Phi(v,\theta,y)}(B)R(\mathbb{X}, dv|\theta, y), \tag{3}$$

for any $B \in \mathcal{B}(\mathcal{P}(\mathbb{X}))$, $C \in \mathcal{B}(\mathbb{Y})$ and $(\theta, y) \in \mathcal{P}(\mathbb{X}) \times \mathbb{Y}$. For notational convenience, let us introduce the real-valued function $H$ (respectively, $h$) defined on $\mathbb{X}\times\mathbb{Y}\times\{0,1\}\times\{0,1\}$ (respectively, $\mathbb{X} \times \mathbb{Y} \times \{0,1\}$) by $H(x,y,z,a) = \mathbf{H}(x,y)I_{\{(z,a)=(0,1)\}}$ (respectively, $h(x,y,z) = \mathbf{H}(x,y)I_{\{z=0\}}$).

Consider the following auxiliary model $\mathcal{M} := \big(\mathbb{S}, \mathbb{A}, Q, H, h\big)$ where

(a) the state space is given by $\mathbb{S} = \mathcal{P}(\mathbb{X}) \times \mathbb{Y} \times \{0,1\}$,

(b) the action space is $\mathbb{A} = \{0,1\}$,

(c) the transition probability function $Q$ is the stochastic kernel on $\mathbb{S}$ given $\mathbb{S} \times \mathbb{A}$ defined by $Q(B \times C \times D|\theta, y, z, a) = S(B \times C|\theta, y)\big[\delta_z(D)I_{\{a=0\}} + \delta_1(D)I_{\{a=1\}}\big]$ for any $B \in \mathcal{B}(\mathcal{P}(\mathbb{X}))$, $C \in \mathcal{B}(\mathbb{Y})$, $D \subset \{0,1\}$ and $(\theta, y, z, a) \in \mathbb{S} \times \mathbb{A}$,

(d) the cost-per-stage is $H(\theta, y, z, a)$ and the terminal cost is $h(\theta, y, z)$ for any $(\theta, y, z, a) \in \mathbb{S} \times \mathbb{A}$ (recalling the slight abuse of notation introduced at the end of Section 1).

The underlying idea is that the filtered trajectory is constructed recursively thanks to the Bayes operator $\Phi$, and the kernel $S$ is the driving kernel of the Markov chain of the filter and observations. The optimal stopping problem is then stated as a sequential decision making problem where at each time step the controller may stop (action $a = 1$) or continue (action $a = 0$). The additional variable $z \in \{0,1\}$ indicates whether the trajectory has already been stopped ($z = 1$) or not ($z = 0$).

Introduce $\Omega = \mathbb{S}^{N_0+1}$, $\mathcal{F}$ its associated product $\sigma$-algebra and the coordinate projections $\Theta_t$ (respectively $Y_t$ and $Z_t$) from $\Omega$ to the set $\mathcal{P}(\mathbb{X})$ (respectively $\mathbb{Y}$ and $\{0,1\}$). Let $\Pi^o$ be the set of all deterministic past dependent control policies $\pi = \{\pi_t\}_{t\in[\![0;N_0-1]\!]}$ where $\pi_0$ is a measurable $\mathbb{A}$-valued function defined on $\mathbb{Y} \times \{0,1\}$ and $\pi_t$ is a measurable $\mathbb{A}$-valued function defined on $(\mathbb{Y} \times \{0,1\} \times \mathbb{A})^t \times \mathbb{Y} \times \{0,1\}$ for $t \in [\![1; N_0]\!]$.

Consider an arbitrary policy $\pi \in \Pi^o$. Define the action process $\{A_t\}_{t\in[\![0;N_0-1]\!]}$ by $A_t = \pi_t(Y_0, Z_0, A_0, \ldots, Y_{t-1}, Z_{t-1}, A_{t-1}, Y_t, Z_t)$ for $t \in [\![1; N_0 - 1]\!]$ and $A_0 = \pi_0(Y_0, Z_0)$. Define $\mathcal{F}_t = \sigma\{\Theta_0, Y_0, Z_0, \ldots, \Theta_t, Y_t, Z_t\}$ for $t \in [\![0; N_0]\!]$. According to [2, 4], for an arbitrary policy $\pi \in \Pi^o$ there exists a probability measure $P^\pi_{(\mathbf{x},\mathbf{y})}$ on $\big(\Omega, \mathcal{F}\big)$ which satisfy

i) $P^\pi_{(\mathbf{x},\mathbf{y})}\big((\Theta_0, Y_0, Z_0) \in B \times C\big) = \delta_{\delta_\mathbf{x}}(B)\delta_\mathbf{y}(C)\delta_0(D)$,

ii) $P^\pi_{(\mathbf{x},\mathbf{y})}\big((\Theta_{t+1}, Y_{t+1}, Z_{t+1}) \in B \times C \times D|\mathcal{F}_t\big) = Q(B \times C \times D|\Theta_t, Y_t, Z_t, A_t)$,

for any $B \in \mathcal{B}(\mathcal{P}(\mathbb{X}))$, $C \in \mathcal{B}(\mathbb{Y})$, $D \subset \{0,1\}$, and $t \in [\![0; N_0 - 1]\!]$.

The expectation under the probability $P^\pi_{(\mathbf{x},\mathbf{y})}$ is denoted by $E^\pi_{(\mathbf{x},\mathbf{y})}$. For a policy $\pi \in \Pi^o$, the performance criterion is given by

$$\boldsymbol{\mathcal{H}}_\mathcal{M}(\mathbf{x}, \mathbf{y}, \pi) \;=\; E^\pi_{(\mathbf{x},\mathbf{y})}\Big[\sum_{t=0}^{N_0-1} H(\Theta_t, Y_t, Z_t, A_t)\Big] + E^\pi_{(\mathbf{x},\mathbf{y})}\big[h(\Theta_{N_0}, Y_{N_0}, Z_{N_0})\big]. \tag{4}$$

The optimization problem consists in maximizing the reward function $\boldsymbol{\mathcal{H}}_\mathcal{M}(\mathbf{x}, \mathbf{y}, \pi)$ over $\Pi^o$ and the corresponding value function is

$$\overline{\boldsymbol{\mathcal{H}}}_\mathcal{M}(\mathbf{x}, \mathbf{y}) = \sup_{\pi\in\Pi^o} \boldsymbol{\mathcal{H}}_\mathcal{M}(\mathbf{x}, \mathbf{y}, \pi). \tag{5}$$

It can be computed using dynamic programming. Consider the Bellman operator $\mathfrak{B}$ defined on $\mathbb{B}(\mathcal{P}(\mathbb{X}) \times \mathbb{Y})$ by

$$\mathfrak{B}f(\theta, y) = \max\{\mathbf{H}(\theta, y), Sf(\theta, y)\}, \tag{6}$$

for $f \in \mathbb{B}(\mathcal{P}(\mathbb{X}) \times \mathbb{Y})$. It should be clear that under Assumption D, $\mathfrak{B}$ maps $\mathbb{L}(\mathcal{P}(\mathbb{X}) \times \mathbb{Y})$ onto $\mathbb{B}(\mathcal{P}(\mathbb{X}) \times \mathbb{Y})$. For notational convenience, $\mathfrak{B}^k$ denotes the $k$-th iteration of $\mathfrak{B}$ recursively defined by $\mathfrak{B}^0 f = f$, $\mathfrak{B}^1 f = \mathfrak{B} f$ and $\mathfrak{B}^k f = \mathfrak{B}(\mathfrak{B}^{k-1} f)$ for $k \in [\![2; N_0]\!]$ and $f \in \mathbb{B}(\mathcal{P}(\mathbb{X}) \times \mathbb{Y})$.

**Theorem 2.2** *Suppose Assumptions (A1), B and D hold. Then*

$$\overline{\boldsymbol{\mathcal{H}}}(\mathbf{x}, \mathbf{y}) = \overline{\boldsymbol{\mathcal{H}}}_{\mathcal{M}}(\mathbf{x}, \mathbf{y}) = \mathfrak{B}^{N_0} \mathbf{H}(\delta_{\mathbf{x}}, \mathbf{y}). \tag{7}$$

**Proof:** See Appendix A. □

# 3 Approximation results

We now build our approximation procedure for the value function $\overline{\boldsymbol{\mathcal{H}}}_{\mathcal{M}}(\mathbf{x}, \mathbf{y})$. It is based on two discretizations. First, we replace the continuous state space $\mathbb{X}$ of the hidden variable by a discrete one $\Gamma_X^N$ with cardinal $N$. Thus, model $\mathcal{M}$ can be approximated by a similar sequential decision making problem for a Markov chain on the finite dimensional state space $\mathcal{P}(\Gamma_X^N) \times \mathbb{Y}$. We then discretize the latter Markov chain using time-dependent grids following the same procedure as in [7]. In both steps, the discretization grids we use are quantization grids. They are especially appealing because they are optimized so that there are more points in the areas of high density, and they allow to control the discretization error in $L_2$-norm as long as the underlying operators have Lipschitz continuity properties. In this section, we first recall the basics of optimal quantization, then present the two discretization steps and derive the discretization error.

## 3.1 Optimal quantization

Consider an $\mathbb{R}^d$-valued random variable $Z$ defined on a probability space $(G, \mathcal{G}, \mathbb{P})$ (with corresponding expectation operator $\mathbb{E}$) such that $\|Z\|_2 < \infty$ where $\|Z\|_2$ denotes the $L_2$-norm of $Z$. Let $N$ be a fixed integer. The optimal $L_2$-quantization of the random variable $Z$ consists in finding the best possible $L_2$-approximation of $Z$ by a random variable $\widehat{Z}_N$ on $(G, \mathcal{G}, \mathbb{P})$ taking at most $N$ values in $\mathbb{R}^d$, which will be denoted by $\{z_N^1, \ldots, z_N^N\}$. The asymptotic properties of the $L_2$-quantization are summarized in the following result (see, e.g., [1, Theorem 3]), which uses the notation $p_\Gamma(z)$ for the closest neighbor projection of $z \in \mathbb{R}^d$ on a grid $\Gamma = \{z_1, \ldots, z_N\} \subseteq \mathbb{R}^d$.

**Theorem 3.1** *Let $Z$ be an $\mathbb{R}^d$-valued random variable on $(G, \mathcal{G}, \mathbb{P})$, and suppose that for some $\epsilon > 0$ we have $\mathbb{E}[|Z|^{2+\epsilon}] < +\infty$. Then*

$$\lim_{N \to \infty} N^{2/d} \min_{|\Gamma| \leq N} \|Z - p_\Gamma(Z)\|_2^2 \quad = \quad J_{d,2} \int_{\mathbb{R}^d} |h_Z(u)|^{d/(2+d)}(u) du,$$

*where $h_Z(u)$ denotes the density of the absolutely continuous part of the distribution of $Z$ with respect to the Lebesgue measure on $\mathbb{R}^d$, and $J_{2,d}$ is a universal constant.*

Finally, let us mention that there exist algorithms that can numerically find, for a fixed $N$, the quantization of $Z$ (or, equivalently, the grid $\{z_N^1, \ldots, z_N^N\}$ attaining the minimum in Theorem 3.1 above and its distribution) as soon as $Z$ is simulatable. Basically, the quantization grids for the variable $Z$ will have more points in the areas of high density, and fewer points in the areas of low density for $Z$.

### 3.2 First approximation

The main originality of our work is to propose a discretization of the state space $\mathbb{X}$ based on the quantization of the reference measure $\lambda$ defined in Assumption (A1). It greatly helps minimizing the computational burden as only one grid is required, instead of a series of grids as one usually does when trying to quantize accurately a Markov chain. In addition, we obtain bounds for the error. However this come at a cost: in order to guarantee that the approximated transition kernel is still a Markov kernel, the density $r$ also appears in the denominator. This is why we need the lower bound of Assumption C to control the error.

To build the approximation and evaluate the error thus entailed, we first quantize the reference probability $\lambda$. Then we replace it by its quantized approximation $\lambda_N$ in the definitions of kernels $R$, $S$, the Bayes operator $\Phi$ and plug these approximations into the Bellman operator. We obtain an approximate Bellman operator $\mathfrak{B}_N$ and our first approximation of the value function is build by iterating $\mathfrak{B}_N$, following Equation (7).

According to the previous discussion of Section 3.1, given an integer $N$, let $\widehat{X}_N$ be the optimal $L_2$-quantization of the random variable $X$ with distribution $\lambda$ on a probability space $(G, \mathcal{G}, \mathbb{P})$ ($\mathbb{E}[\cdot]$ will stand for the expectation associated to $\mathbb{P}$). Let us denote by $\Gamma_X^N = \{x_N^1, \dots, x_N^N\}$ an optimal grid. There is no loss of generality to assume that $x_N^1 = \mathbf{x}$. We write $\lambda_N$ for the distribution of $\widehat{X}_N$, that is, $\lambda_N(du) = \sum_{i=1}^N \mathbb{P}(\widehat{X}_N = x_N^i)\delta_{x_N^i}(du)$ and

$$\epsilon_N = \|X - \widehat{X}_N\|_2$$

for the $L_2$-quantization error between $X$ and $\widehat{X}_N$. Assume also the existence of a random variable $Y$ with distribution $\nu$ on $(G, \mathcal{G}, \mathbb{P})$.

We define the quantized approximations of kernels $R$ and $S$ by plugging-in $\lambda_N$ as follows. Consider the stochastic kernel $R_N$ on $\mathbb{X} \times \mathbb{Y}$ where

$$R_N(B \times C|x, y) = \int_{B\times C} \frac{r(u, v, x, y)}{r(\lambda_N, \nu, x, y)}\lambda_N(du)\nu(dv), \tag{8}$$

for any $(x, y) \in \mathbb{X} \times \mathbb{Y}$, $B \in \mathcal{B}(\mathbb{X})$, $C \in \mathcal{B}(\mathbb{Y})$. Note that the support of $R_N(\cdot|x, y)$ is actually $\Gamma_X^N \times \mathbb{Y}$. Let us introduce the stochastic kernel $S_N$ on $\mathcal{P}(\mathbb{X}) \times \mathbb{Y}$ defined by

$$S_N(B \times C|\theta, y) = \int_C \delta_{\Phi_N(v,\theta,y)}(B)R_N(\mathbb{X}, dv|\theta, y), \tag{9}$$

where $\Phi_N : \mathbb{Y} \times \mathcal{P}(\mathbb{X}) \times \mathbb{Y} \to \mathcal{P}(\mathbb{X})$ given by

$$\Phi_N(v, \theta, y)(du) = \frac{r(u, v, \theta, y)}{r(\lambda_N, v, \theta, y)}\lambda_N(du). \tag{10}$$

Here again, $\Phi_N$ actually maps $\mathbb{Y} \times \mathcal{P}(\mathbb{X}) \times \mathbb{Y}$ onto $\mathcal{P}(\Gamma_X^N)$ and the support of $S_N(\cdot|\theta, y)$ is $\mathcal{P}(\Gamma_X^N) \times \mathbb{Y}$. Next we define the approximated Bellman operator. Consider the operator $\mathfrak{B}_N$ defined on $\mathbb{B}(\mathcal{P}(\mathbb{X}) \times \mathbb{Y})$ by

$$\mathfrak{B}_N f(\theta, y) = \max\{\mathbf{H}(\theta, y), S_N f(\theta, y)\} \tag{11}$$

for $f \in \mathbb{B}(\mathcal{P}(\mathbb{X}) \times \mathbb{Y})$. It should be clear that under Assumption D, $\mathfrak{B}_N$ maps $\mathbb{B}(\mathcal{P}(\mathbb{X}) \times \mathbb{Y})$ onto $\mathbb{B}(\mathcal{P}(\mathbb{X}) \times \mathbb{Y})$ and $\mathbb{B}(\mathcal{P}(\Gamma_X^N) \times \mathbb{Y})$ onto $\mathbb{B}(\mathcal{P}(\Gamma_X^N) \times \mathbb{Y})$. For notational convenience, $\mathfrak{B}_N^k$ denotes the

$k$-th iteration of $\mathfrak{B}_N$. The first approximate value function is then defined as the $N_0$-iterate of $\mathbb{B}_N$ on $\mathbf{H}$:

$$\overline{\boldsymbol{\mathcal{H}}}_{\mathcal{M},N}(\mathbf{x},\mathbf{y}) = \mathfrak{B}_N^{N_0}\mathbf{H}(\delta_{\mathbf{x}},\mathbf{y}). \tag{12}$$

We now study the error induced by this discretization on the value function. To do so, we study how the quantization error is propagated through the Bellman operator. We start with technical results on the density of the Markov chain integrated with respect to $\lambda$ and with respect to its quantized approximation $\lambda_N$.

**Lemma 3.2** *For any $(\theta, y) \in \mathcal{P}(\mathbb{X}) \times \mathbb{Y}$ and $N \in \mathbb{N}$ such that $\epsilon_N \leq \dfrac{1}{2L_r}$, one has*

$$\frac{1}{r(\lambda_N,\nu,\theta,y)} \leq 2 \quad \text{and} \quad \Big|1 - \frac{1}{r(\lambda_N,\nu,\theta,y)}\Big| \leq 2L_r\epsilon_N.$$

**Proof:** Clearly, we have $\big|1 - r(\lambda_N,\nu,\theta,y)\big| = \big|r(\lambda,\nu,\theta,y) - r(\lambda_N,\nu,\theta,y)\big| \leq L_r\epsilon_N$ by using Assumption (B2) and so, $\dfrac{1}{r(\lambda_N,\nu,\theta,y)} \leq 2$ for $N \in \mathbb{N}$ satisfying $\epsilon_N \leq \dfrac{1}{2L_r}$. Therefore,

$$\Big|1 - \frac{1}{r(\lambda_N,\nu,\theta,y)}\Big| \leq \frac{1}{r(\lambda_N,\nu,\theta,y)}\big|r(\lambda,\nu,\theta,y) - r(\lambda_N,\nu,\theta,y)\big| \leq \frac{1}{r(\lambda_N,\nu,\theta,y)}L_r\epsilon_N,$$

giving the result. □

**Lemma 3.3** *Suppose Assumption C holds. For any $(\theta, y) \in \mathcal{P}(\mathbb{X}) \times \mathbb{Y}$ and $N \in \mathbb{N}$ such that $\epsilon_N \leq \dfrac{1}{2\delta L_r}$, one has $\dfrac{1}{r(\lambda_N,v,\theta,y)} \leq 2\delta$.*

**Proof:** The proof is similar to the one of Lemma 3.2 and is therefore, omitted. □

**Lemma 3.4** *Suppose Assumptions B and C hold. For any $\theta \in \mathcal{P}(\mathbb{X})$ and $N \in \mathbb{N}$ such that $\epsilon_N \leq \frac{1}{2\delta L_r}$, one has*

$$\sup_{y\in\mathbb{Y}} d_{\mathcal{P}}\big(\Phi(Y,\theta,y), \Phi_N(Y,\theta,y)\big) \leq \delta\big[(2\delta+1)L_r + \overline{r}\big]\epsilon_N.$$

**Proof:** Consider $f \in \mathbb{L}_1(\mathbb{X})$. We have

$$\begin{aligned}
\Big|\int_{\mathbb{X}} f(x)\Phi(v,\theta,y)(dx) - \int_{\mathbb{X}} f(x)\Phi_N(v,\theta,y)(dx)\Big| &\leq \mathbb{E}\Big[\Big|f(X)\frac{r(X,v,\theta,y)}{r(\lambda,v,\theta,y)} - f(X_N)\frac{r(X_N,v,\theta,y)}{r(\lambda_N,v,\theta,y)}\Big|\Big] \\
&\leq \mathbb{E}\Big[|f(X_N)|r(X_N,v,\theta,y)\Big|\frac{1}{r(\lambda,v,\theta,y)} - \frac{1}{r(\lambda_N,v,\theta,y)}\Big|\Big] \\
&\quad + \mathbb{E}\Big[\frac{|f(X)|}{r(\lambda,v,\theta,y)}\big|r(X,v,\theta,y) - r(X_N,v,\theta,y)\big|\Big] \\
&\quad + \mathbb{E}\Big[\frac{r(X_N,v,\theta,y)}{r(\lambda,v,\theta,y)}|f(X) - f(X_N)|\Big].
\end{aligned}$$

By using Assumptions (B1) and C, it follows that

$$\begin{aligned}
&\Big|\int_{\mathbb{X}} f(x)\Phi(v,\theta,y)(dx) - \int_{\mathbb{X}} f(x)\Phi_N(v,\theta,y)(dx)\Big| \\
&\qquad \leq \overline{r}\mathbb{E}\Big[\Big|\frac{1}{r(\lambda,v,\theta,y)} - \frac{1}{r(\lambda_N,v,\theta,y)}\Big|\Big] + \delta L_r\epsilon_N + \delta\overline{r}\epsilon_N.
\end{aligned} \tag{13}$$

However, for $N \in \mathbb{N}$ satisfying $\epsilon_N \leq \frac{1}{2\delta L_r}$ we get from Lemma 3.3 that

$$\Big|\frac{1}{r(\lambda,v,\theta,y)} - \frac{1}{r(\lambda_N,v,\theta,y)}\Big| \leq \frac{1}{r(\lambda,v,\theta,y)r(\lambda_N,v,\theta,y)} L_r\epsilon_N \leq 2\delta^2 L_r \epsilon_N$$

and with equation (13), this shows the result. □

We now need to ensure that both $\mathfrak{B}$ and $\mathfrak{B}_N$ operate on $\mathbb{L}(\mathcal{P}(\mathbb{X}) \times \mathbb{Y})$.

**Lemma 3.5** *Suppose Assumptions B, C and D hold. For any $\theta, \theta' \in \mathcal{P}(\mathbb{X})$, $y, y' \in \mathbf{Y}$, $N \in \mathbb{N}$ and $f \in \mathbb{L}(\mathcal{P}(\mathbb{X}) \times \mathbb{Y})$ one has $\mathfrak{B}f \in \mathbb{L}(\mathcal{P}(\mathbb{X}) \times \mathbb{Y})$.*

**Proof:** Consider $f \in \mathbb{L}(\mathcal{P}(\mathbb{X}) \times \mathbb{Y})$ and $(\theta, y), (\theta', y') \in \mathcal{P}(\mathbb{X}) \times \mathbb{Y}$. On the one hand,

$$\begin{aligned}
|\mathfrak{B}f(\theta,y)| &\leq \max\{|\mathbf{H}(\theta,y)|; |Sf(\theta,y)|\} \\
&\leq \max\left\{\|\mathbf{H}\|_{\mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})}; \int_{\mathbb{X}\times\mathbb{Y}} \big|f\big(\Phi(y',\theta,y),y'\big)\big|\, r(x',y',\theta,y)\lambda(dx')\nu(dy')\right\} \\
&\leq \max\left\{\|\mathbf{H}\|_{\mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})}; \|f\|_{\mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})}\right\}. \qquad (14)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
|Sf(\theta,y) - Sf(\theta',y')| &\leq \mathbb{E}\left[\big|f\big(\Phi(Y,\theta,y),Y\big)\big|\,\big|r(X,Y,\theta,y) - r(X,Y,\theta',y')\big|\right] \\
&\quad + \mathbb{E}\left[r(X,Y,\theta',y')\,\big|f\big(\Phi(Y,\theta,y),Y\big) - f\big(\Phi(Y,\theta',y'),Y\big)\big|\right] \\
&\leq \|f\|_{\mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})}(\overline{r} + L_r)[d_{\mathcal{P}}(\theta,\theta') + |y-y'|] \\
&\quad + \overline{r}\|f\|_{\mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})}\mathbb{E}\left[d_{\mathcal{P}}(\Phi(Y,\theta,y),\Phi(Y,\theta',y'))\right]. \qquad (15)
\end{aligned}$$

Let now $g \in \mathbb{L}_1(\mathcal{P}(\mathbb{X}) \times \mathbb{Y})$ and $v \in \mathbb{Y}$. Then, one has

$$\begin{aligned}
&\left|\int_{\mathbb{X}} g(u)\Phi(v,\theta,y)(du) - \int_{\mathbb{X}} g(u)\Phi(v,\theta',y')(dy)\right| \\
&\leq \mathbb{E}\left[|g(X)|\left|\frac{r(X,v,\theta,y)}{r(\lambda,v,\theta,y)} - \frac{r(X,v,\theta',y')}{r(\lambda,v,\theta',y')}\right|\right] \\
&\leq \mathbb{E}\left[\frac{1}{r(\lambda,v,\theta,y)}|r(X,v,\theta,y) - r(X,v,\theta',y')|\right] \\
&\quad + \mathbb{E}\left[r(X,v,\theta',y')\left|\frac{1}{r(\lambda,v,\theta,y)} - \frac{1}{r(\lambda,v,\theta',y')}\right|\right] \\
&\leq \delta(\overline{r} + L_r)(1+\delta\overline{r})[d_{\mathcal{P}}(\theta,\theta') + |y-y'|] \qquad (16)
\end{aligned}$$

by using assumptions B and C. So, one has

$$\mathbb{E}\left[d_{\mathcal{P}}(\Phi(Y,\theta,y),\Phi(Y,\theta',y'))\right] \leq \delta(\overline{r} + L_r)(1+\delta\overline{r})[d_{\mathcal{P}}(\theta,\theta') + |y-y'|]. \qquad (17)$$

Then, by using assumption D, it is straightforward to write

$$\begin{aligned}
|\mathfrak{B}f(\theta,y) - \mathfrak{B}f(\theta',y')| &\leq |\mathbf{H}(\theta,y) - \mathbf{H}(\theta',y')| + |Sf(\theta,y) - Sf(\theta',y')| \\
&\leq L_{\mathfrak{B}f}[d_{\mathcal{P}}(\theta,\theta') + |y-y'|] \qquad (18)
\end{aligned}$$

with

$$L_{\mathfrak{B}f} = \|\mathbf{H}\|_{\mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})} + \|f\|_{\mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})}(\overline{r} + L_r)\big(1 + \delta\overline{r}(1+\overline{r}\delta)\big). \qquad (19)$$

Thus, $\mathfrak{B}f$ is bounded and Lipschitz-continuous on $\mathcal{P}(\mathbb{X}) \times \mathbb{Y}$.

□

**Proposition 3.6** *Suppose Assumptions A, B and C hold. Let $N \in \mathbb{N}$ satisfying $\epsilon_N \leq \frac{1}{2L_r}(1 \wedge \frac{1}{\delta})$,*

$$\big|\mathfrak{B}f(\theta,y) - \mathfrak{B}_N f(\theta,y)\big| \leq \|f\|_{\mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})} K_1 \epsilon_N \tag{20}$$

*for any $(\theta,y) \in \mathcal{P}(\mathbb{X})\times\mathbb{Y}$ and $f \in \mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})$ with*

$$K_1 = L_r(1+2\overline{r}L_r) + \delta\overline{r}\big[\overline{r} + L_r(2+2\delta\overline{r})\big]. \tag{21}$$

**Proof:** Consider $(\theta,y) \in \mathcal{P}(\mathbb{X})\times\mathbb{Y}$ and $f \in \mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})$. Clearly, we have

$$\big|\mathfrak{B}f(\theta,y) - \mathfrak{B}_N f(\theta,y)\big| \leq \big|Sf(\theta,y) - S_N f(\theta,y)\big|.$$

By using the definition of $S$ and $S_N$ (see equations (3) and (9) respectively), we have

$$\begin{aligned}
\big|Sf(\theta,y) - S_N f(\theta,y)\big| &\leq \mathbb{E}\bigg[\Big|f(\Phi(Y,\theta,y),Y)r(\lambda,Y,\theta,y) - f(\Phi_N(Y,\theta,y),Y)\frac{r(\lambda_N,Y,\theta,y)}{r(\lambda_N,\nu,\theta,y)}\Big|\bigg] \\
&\leq \mathbb{E}\Big[r(\lambda,Y,\theta,y)\big|f(\Phi(Y,\theta,y),Y) - f(\Phi_N(Y,\theta,y),Y)\big|\Big] \\
&\quad + \mathbb{E}\Big[\big|f(\Phi_N(Y,\theta,y),Y)\big|\big|r(\lambda,Y,\theta,y) - r(\lambda_N,Y,\theta,y)\big|\Big] \\
&\quad + \mathbb{E}\bigg[r(\lambda_N,Y,\theta,y)\big|f(\Phi_N(Y,\theta,y),Y)\big|\Big|1 - \frac{1}{r(\lambda_N,\nu,\theta,y)}\Big|\bigg].
\end{aligned}$$

Consequently, it follows that

$$\begin{aligned}
\big|\mathfrak{B}f(\theta,y) - \mathfrak{B}_N f(\theta,y)\big| &\leq \overline{r}\|f\|_{\mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})}\mathbb{E}\Big[d_{\mathcal{P}}\big(\Phi(Y,\theta,y),\Phi_N(Y,\theta,y)\big)\Big] \\
&\quad + \|f\|_{\mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})}L_r\epsilon_N + \overline{r}\|f\|_{\mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})}\Big|1 - \frac{1}{r(\lambda_N,\nu,\theta,y)}\Big|.
\end{aligned}$$

By using Lemma 3.4 and 3.2, we get the result. □

We can now state and prove the main result of this section bounding the error between the true value function and its quantized approximation.

**Theorem 3.7** *Suppose Assumptions A, B, C and D hold. Let $N \in \mathbb{N}$ satisfying $\epsilon_N \leq \frac{1}{2L_r}(1 \wedge \frac{1}{\delta})$. Then, one has*

$$\big|\overline{\boldsymbol{\mathcal{H}}}_{\mathcal{M}}(\mathbf{x},\mathbf{y}) - \overline{\boldsymbol{\mathcal{H}}}_{\mathcal{M},N}(\mathbf{x},\mathbf{y})\big| \leq K_1\epsilon_N \sum_{k=0}^{N_0-1} \|\mathfrak{B}^k\mathbf{H}\|_{\mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})}. \tag{22}$$

**Proof:** First let us show by induction that

$$\big|\mathfrak{B}^k f(\theta,y) - \mathfrak{B}_N^k f(\theta,y)\big| \leq K_1\epsilon_N \sum_{j=0}^{k-1} \|\mathfrak{B}^j f\|_{\mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})} \tag{23}$$

for any $(\theta,y) \in \mathcal{P}(\mathbb{X})\times\mathbb{Y}$, $f \in \mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})$ and $k \in [\![1; N_0]\!]$. From Proposition 3.6, the claim is true for $k=1$. Now, assume that equation (23) holds for $k \in [\![1; N_0-1]\!]$. Then,

$$\begin{aligned}
\big|\mathfrak{B}^{k+1} f(\theta,y) - \mathfrak{B}_N^{k+1} f(\theta,y)\big| &\leq \big|\mathfrak{B}(\mathfrak{B}^k f)(\theta,y) - \mathfrak{B}_N(\mathfrak{B}^k f)(\theta,y)\big| \\
&\quad + \big|\mathfrak{B}_N(\mathfrak{B}^k f)(\theta,y) - \mathfrak{B}_N(\mathfrak{B}_N^k f)(\theta,y)\big|.
\end{aligned} \tag{24}$$

From equation (20), Lemma 3.5 and recalling the definition of $\mathfrak{B}_N$ (see equation (11)) we get

$$\left|\mathfrak{B}^{k+1}f(\theta,y)-\mathfrak{B}_N^{k+1}f(\theta,y)\right| \leq \|\mathfrak{B}^k f\|_{\mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})}K_1\epsilon_N + \left|S_N(\mathfrak{B}^k f)(\theta,y)-S_N(\mathfrak{B}_N^k f)(\theta,y)\right|. \quad (25)$$

Now, combining (9) and the induction hypothesis we have

$$\begin{aligned}\left|S_N(\mathfrak{B}^k f)(\theta,y)-S_N(\mathfrak{B}_N^k f)(\theta,y)\right| &\leq \int_{\mathbb{Y}}\left|\mathfrak{B}^k f(\Phi_N(v,\theta,y),v)-\mathfrak{B}_N^k f(\Phi_N(v,\theta,y),v)\right| R_N(\mathbb{X},dv|\theta,y)\\ &\leq K_1\epsilon_N\sum_{j=0}^{k-1}\|\mathfrak{B}^j f\|_{\mathbb{L}(\mathcal{P}(\mathbb{X})}, \end{aligned} \quad (26)$$

and so from equations (24)-(26) we obtain that (23) holds for any $k \in [\![1; N_0]\!]$. Finally, recalling that $\overline{\boldsymbol{\mathcal{H}}}_{\mathcal{M}}(\mathbf{x},\mathbf{y}) = \mathfrak{B}^{N_0}\mathbf{H}(\delta_{\mathbf{x}},\mathbf{y})$ and $\overline{\boldsymbol{\mathcal{H}}}_{\mathcal{M},N}(\mathbf{x},\mathbf{y}) = \mathfrak{B}_N^{N_0}\mathbf{H}(\delta_{\mathbf{x}},\mathbf{y})$, we obtain the result by applying (23) to $f = \mathbf{H}$ with $k = N_0$ since $\mathbf{H} \in \mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})$ by Assumption D. □

### 3.3 Second approximation

The approximate value function $\overline{\boldsymbol{\mathcal{H}}}_{\mathcal{M},N}(\mathbf{x},\mathbf{y}) = \mathfrak{B}_N^{N_0}\mathbf{H}(\delta_{\mathbf{x}},\mathbf{y})$ is not directly computable as it involves a recursion of functions defined on the continuous space $\mathcal{P}(\Gamma_X^N)\times\mathbb{Y}$. In order to obtain a numerically tractable recursion, one additional discretization procedure is required. We first introduce the Markov chain $\Psi^N$ with transition kernel $S_N$, then rewrite the iteration of the Bellman operators $\mathfrak{B}_N$ in terms of conditional expectations involving this chain, and finally propose and approximation of the latter conditional expectations based on the quantization of the chain $\Psi^N$. Following the idea of [7], instead of discretizing the two coordinates (filter and observations) separately, we discretize them jointly exploiting the Markov property of $\Psi^N$.

Let us denote by $\{\Psi_t^N\}_{t\in[\![0;N_0]\!]}$ the Markov chain with transition kernel $S_N$ and initial distribution $(\delta_{\mathbf{x}},\mathbf{y})$. By definition of $S_N$ we have that $S_N(\mathcal{P}(\Gamma_X^N)\times\mathbb{Y}|\theta,y) = 1$ for any $(\theta,y)\in\mathcal{P}(\mathbb{X})\times\mathbb{Y}$ and that $\delta_{\mathbf{x}}\in\mathcal{P}(\Gamma_X^N)$. Moreover, it is clear that $\mathcal{P}(\Gamma_X^N)$ can be identified with the $N$-simplex in $\mathbb{R}^N$ denoted by $\mathbb{S}^N$. Therefore, by a slight abuse of notation we will consider from now on that the state space of the Markov chain $\{\Psi_t^N\}_{t\in[\![0;N_0]\!]}$ is given by $\mathbb{S}^N\times\mathbb{Y}\subset\mathbb{R}^{N+n}$. For notational convenience, the stochastic kernel associated with $\{\Psi_t^N\}_{t\in[\![0;N_0]\!]}$ will still be denoted by $S_N$. Thus, our aim is now to rewrite the Bellman operator $\mathfrak{B}_N$ in terms of conditional expectations involving $\{\Psi_t^N\}_{t\in[\![0;N_0]\!]}$, discretize this Markov chain using optimal quantization and see how the approximation error is propagated through the dynamic programming recursion.

First, we rewrite the dynamic programming recursion on functions (12) as a recursion involving conditional expectations. By a slight abuse of notation, we write

$$\mathbf{H}(\psi) = \sum_{j=1}^N \gamma_j \mathbf{H}(x_N^j, y), \quad (27)$$

for $\psi = (\gamma, y)\in\mathbb{S}^N\times\mathbb{Y}$. Define recursively the sequence of real-valued functions $\{V_t^N\}_{t\in[\![0;N_0]\!]}$ on $\mathbb{S}^N\times\mathbb{Y}$ by

$$V_t^N(\psi) = \max\left\{\mathbf{H}(\psi), \mathbb{E}\left[V_{t+1}(\Psi_{t+1}^N)\middle|\Psi_t^N=\psi\right]\right\} \quad (28)$$

for $t\in[\![0;N_0-1]\!]$ and $V_{N_0}^N(\psi) = \mathbf{H}(\psi)$ for $\psi\in\mathbb{S}^N\times\mathbb{Y}$. Note that these dynamic programming equations now go backward in time, with an initialisation at the terminal time $N_0$. By definition of

the operator $\mathfrak{B}_N$ (see equation 11), we have clearly $V_0^N(\Psi_0^N) = \mathfrak{B}_N^{N_0}\mathbf{H}(\delta_{\mathbf{x}}, \mathbf{y})$ and so by Theorem 3.7, $V_0^N(\Psi_0^N) = \overline{\boldsymbol{\mathcal{H}}}_{\mathcal{M},N}(\mathbf{x}, \mathbf{y})$. Thus one just needs to build a numerically computable approximation of function $V_0^N$.

Let $\{\widehat{\Psi}_t^{N,M} = (\widehat{\Theta}_t^{N,M}, \widehat{Y}_t^{N,M})\}_{n\in[\![0;N_0]\!]}$ be the quantization approximation of $\{\Psi_t^N\}_{t\in[\![0;N_0]\!]}$ defined on a probability space $(\overline{G}, \overline{\mathcal{G}}, \overline{\mathbb{P}})$ ($\overline{\mathbb{E}}[\cdot]$ will stand for the expectation associated to $\overline{\mathbb{P}}$). There are several methods to get the quantization of a Markov chain such as the marginal quantization or Markovian quantization approaches. These techniques are roughly speaking based upon the quantization of a random variable as described in section 3.1. We do not want to go into the details of these different approaches. A rather complete exposition of this subject can be found in [1, 5]. We write $\Gamma^M_{\Psi_t^N}$ for the grid of $M$ points used to quantize $\Psi_t^N$ and $\|\Psi_t^N - \widehat{\Psi}_t^{N,M}\|_2$ for the $L_2$-quantization error between $\Psi_t^N$ and $\widehat{\Psi}_t^{N,M}$ under $\overline{\mathbb{P}}$. Define recursively the sequence of real-valued functions $\{\widehat{V}_t^{N,M}\}_{t\in[\![0;N_0]\!]}$ by

$$\widehat{V}_t^{N,M}(\widehat{\psi}) = \max\Big\{\mathbf{H}(\widehat{\psi}), \overline{\mathbb{E}}\big[\widehat{V}_{t+1}^{N,M}(\widehat{\Psi}_{t+1}^N)\big|\widehat{\Psi}_t^N = \widehat{\psi}\big]\Big\},$$

for any $\widehat{\psi} \in \Gamma^M_{\Psi_t^N}$, $t \in [\![0; N_0 - 1]\!]$ and $\widehat{V}_{N_0}^{N,M}(\widehat{\psi}) = \mathbf{H}(\widehat{\psi})$ for $\widehat{\psi} \in \Gamma^M_{\Psi_{N_0}^N}$. As $\{\widehat{\Psi}_t^{N,M}\}$ is now a (inhomogeneous) Markov chain on a finite state space, the conditional expectations above are just weighted sums and can be computed numerically. Before stating the main result of this section regarding the convergence of $\widehat{V}_t^{N,M}$ to $\widehat{V}_t^N$, we need additional technical results on the Lipschitz regularity of $V^N$ and $V^{N,M}$.

**Lemma 3.8** *Suppose Assumptions A, B, C and D hold. Let $N \in \mathbb{N}$ satisfying $\epsilon_N \leq \frac{1}{2L_r}(1 \wedge \frac{1}{\delta})$. Then $V_t^N \in \mathbb{L}(\mathbb{S}^N \times \mathbb{Y})$ and $\|V_t^N\|_{sup} \leq \|\mathbf{H}\|_{sup}$ for $t \in [\![0; N_0]\!]$. Moreover, one has*

$$L_{V_t^N} \leq 4\sqrt{N}\Big[(1 + 2\overline{r})\|\mathbf{H}\|_{sup} + 2\overline{r}\delta(1 + 2\overline{r}\delta)L_{V_{t+1}^N}\Big](\overline{r} + L_r) + 2\sqrt{N}(\|\mathbf{H}\|_{sup} + L_{\mathbf{H}}) \qquad (29)$$

*for $t \in [\![0; N_0 - 1]\!]$ and $L_{V_{N_0}^N} \leq 2\sqrt{N}(\|\mathbf{H}\|_{sup} + L_{\mathbf{H}})$.*

**Proof:** According to equation (27), it is clear that $\|V_{N_0}^N\|_{sup} \leq \|\mathbf{H}\|_{sup} = \sup_{(x,y)\in\mathbb{X}\times\mathbb{Y}} |\mathbf{H}(x, y)|$. Moreover, for $\psi = (\gamma, y)$ and $\psi' = (\gamma', y')$ in $\mathbb{S}^N \times \mathbb{Y}$

$$\begin{aligned}|V_{N_0}^N(\psi) - V_{N_0}^N(\psi')| &\leq (\|\mathbf{H}\|_{sup} + L_{\mathbf{H}})\Big[\sum_{j=1}^N |\gamma_j - \gamma'_j| + \big|y - y'\big|\Big] \\ &\leq \|\mathbf{H}\|_{\mathbb{L}(\mathbb{S}^N\times\mathbb{Y})}\Big[\sqrt{N}|\gamma - \gamma'| + \big|y - y'\big|\Big] \leq 2\sqrt{N}\|\mathbf{H}\|_{\mathbb{L}(\mathbb{S}^N\times\mathbb{Y})}|\psi - \psi'|,\end{aligned}$$

giving the Lipschitz constant of $V_{N_0}^N$.
Now, by a slight abuse of notation, $\Phi_N(Y, \gamma, y)$ is identified with the vector in $\mathbb{S}^N$ which the $j^{th}$ component is given by $\Phi_N(Y, \gamma, y)(x_j^N)$ and $r(X_N, Y, \gamma, y)$ (respectively, $r(\lambda_N, \nu, \gamma, y)$) denotes $\sum_{j=1}^N \gamma_j r(X_N, Y, x_j^N, y)$ (respectively, $\sum_{j=1}^N \gamma_j r(\lambda_N, \nu, x_j^N, y)$).

Consider $g \in \mathbb{L}(\mathbb{S}^N \times \mathbb{Y})$ and $\psi = (\gamma, y)$, $\psi' = (\gamma', y')$ in $\mathbb{S}^N \times \mathbb{Y}$. For any $t \in [\![0; N_0 - 1]\!]$, we have

$$\begin{aligned}
\Big|\mathbb{E}\big[g(\Psi^N_{t+1})\big|\Psi^N_t = \psi\big] - \mathbb{E}^N\big[g(\Psi^N_{t+1})\big|\Psi^N_t = \psi\big]\Big| &= \big|S_N g(\gamma, y) - S_N g(\gamma', y')\big| \\
&\leq \mathbb{E}\Big[\Big|g(\Phi_N(Y,\gamma,y),Y)\frac{r(\lambda_N, Y, \gamma, y)}{r(\lambda_N, \nu, \gamma, y)} - g(\Phi_N(Y,\gamma',y'),Y)\frac{r(\lambda_N, Y, \gamma', y')}{r(\lambda_N, \nu, \gamma', y')}\Big|\Big] \\
&\leq \mathbb{E}\Big[\frac{|g(\Phi_N(Y,\gamma,y),Y)|}{r(\lambda_N, \nu, \gamma, y)}\big|r(\lambda_N, \nu, \gamma, y) - r(\lambda_N, \nu, \gamma', y')\big|\Big] \\
&\quad + \mathbb{E}\Big[|g(\Phi_N(Y,\gamma,y),Y)|r(\lambda_N, \nu, \gamma', y')\Big|\frac{1}{r(\lambda_N, \nu, \gamma, y)} - \frac{1}{r(\lambda_N, \nu, \gamma', y')}\Big|\Big] \\
&\quad + \mathbb{E}\Big[\frac{r(\lambda_N, Y, \gamma', y')}{r(\lambda_N, \nu, \gamma', y')}|g(\Phi_N(Y,\gamma,y),Y) - g(\Phi_N(Y,\gamma',y'),Y)|\Big]. \qquad (30)
\end{aligned}$$

By using Lemma 3.2 and Assumption B we have

$$\begin{aligned}
&\mathbb{E}\Big[\frac{|g(\Phi_N(Y,\gamma,y),Y)|}{r(\lambda_N, \nu, \gamma, y)}\big|r(X_N, \nu, \gamma, y) - r(X_N, \nu, \gamma', y')\big|\Big] \\
&\quad \leq 2\|g\|_{sup}(\overline{r} + L_r)\Big[\sum_{j=1}^N |\gamma_j - \gamma'_j| + \big|y - y'\big|\Big] \leq 2\|g\|_{sup}(\overline{r} + L_r)\Big[\sqrt{N}|\gamma - \gamma'| + \big|y - y'\big|\Big]. \qquad (31)
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathbb{E}\Big[|g(\Phi_N(Y,\gamma,y),Y)|r(\lambda_N, \nu, \gamma', y')\Big|\frac{1}{r(\lambda_N, \nu, \gamma, y)} - \frac{1}{r(\lambda_N, \nu, \gamma', y')}\Big|\Big] \\
&\quad \leq \mathbb{E}\Big[|g(\Phi_N(Y,\gamma,y),Y)|\frac{r(\lambda_N, \nu, \gamma', y')}{r(\lambda_N, \nu, \gamma, y)r(\lambda_N, \nu, \gamma', y')}\Big|r(\lambda_N, \nu, \gamma, y) - r(\lambda_N, \nu, \gamma', y')\Big|\Big] \\
&\quad \leq 4\|g\|_{sup}\overline{r}(\overline{r} + L_r)\Big[\sqrt{N}|\gamma - \gamma'| + \big|y - y'\big|\Big], \qquad (32)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\Big[\frac{r(\lambda_N, Y, \gamma', y')}{r(\lambda_N, \nu, \gamma', y')}|g(\Phi_N(Y,\gamma,y),Y) - g(\Phi_N(Y,\gamma',y'),Y)|\Big] \\
&\qquad\qquad \leq 2\overline{r}L_g\mathbb{E}\big[\big|\Phi_N(Y,\gamma,y) - \Phi_N(Y,\gamma',y')\big|\big]. \qquad (33)
\end{aligned}$$

Moreover, from the definition of the discrete measure $\Phi_N$ (see equation (10))

$$\begin{aligned}
\mathbb{E}\big[\big|\Phi_N(Y,\gamma,y) - \Phi_N(Y,\gamma',y')\big|\big] &\leq \mathbb{E}\Big[\Big(\sum_{j=1}^N \lambda_N(x^j_N)^2\Big|\frac{r(x^j_N, Y, \gamma, y)}{r(\lambda_N, Y, \gamma, y)} - \frac{r(x^j_N, Y, \gamma', y')}{r(\lambda_N, Y, \gamma', y')}\Big|^2\Big)^{1/2}\Big] \\
&\leq \sup_{(u,v)\in\Gamma^N_X\times\mathbb{Y}}\Big|\frac{r(u, v, \gamma, y)}{r(\lambda_N, v, \gamma, y)} - \frac{r(u, v, \gamma', y')}{r(\lambda_N, v, \gamma', y')}\Big| \\
&\leq \sup_{(u,v)\in\Gamma^N_X\times\mathbb{Y}}\Big[\frac{1}{r(\lambda_N, v, \gamma, y)}\Big|r(u, v, \gamma, y) - r(u, v, \gamma', y')\Big|\Big] \\
&\quad + \sup_{(u,v)\in\Gamma^N_X\times\mathbb{Y}}\Big[r(u, v, \gamma', y')\Big|\frac{1}{r(\lambda_N, v, \gamma, y)} - \frac{1}{r(\lambda_N, v, \gamma', y')}\Big|\Big],
\end{aligned}$$

and so, from Lemma 3.3

$$\begin{aligned}
\mathbb{E}\big[\big|\Phi_N(Y,\gamma,y) - \Phi_N(Y,\gamma',y')\big|\big] &\leq 2\delta(1 + 2\overline{r}\delta)\sup_{(u,v)\in\Gamma^N_X\times\mathbb{Y}}\Big|r(u, v, \theta, y) - r(u, v, \theta', y')\Big| \\
&\leq 2\delta(1 + 2\overline{r}\delta)(\overline{r} + L_r)\Big[\sqrt{N}|\gamma - \gamma'| + \big|y - y'\big|\Big]. \qquad (34)
\end{aligned}$$

Combining equations (30)-(34), we obtain

$$\begin{aligned}\Big|\mathbb{E}\big[g(\Psi_{t+1}^N)\big|\Psi_t^N=\psi\big]-\mathbb{E}^N\big[g(\Psi_{t+1}^N)\big|\Psi_t^N=\psi\big]\Big| &= \big|S_N g(\gamma,y)-S_N g(\gamma',y')\big| \\ &\leq 4\sqrt{N}\Big[(1+2\overline{r})\|g\|_{sup}+2\overline{r}\delta(1+2\overline{r}\delta)L_g\Big](\overline{r}+L_r)\big|\psi-\psi'\big|.\end{aligned}$$

Finally, by using the definition of $V_t^N$ (see equation (28)), we get (29) showing the result. □

We now state and prove the main result of this section.

**Theorem 3.9** *Suppose Assumptions A, B, C and D hold. Let $N\in\mathbb{N}$ satisfying $\epsilon_N\leq\frac{1}{2L_r}(1\wedge\frac{1}{\delta})$. Then*

$$\big|\overline{\boldsymbol{\mathcal{H}}}_{\mathcal{M},N}(\mathbf{x},\mathbf{y})-\widehat{V}_0^{N,M}(\widehat{\Psi}_0^{N,M})\big|\leq\sum_{t=0}^{N_0}L_{V_t^N}\|\Psi_t^N-\widehat{\Psi}_t^{N,M}\|_2.$$

**Proof:** The proof of this result is based on Theorem 2 in [1]. The main difference is that in our setting, the transition kernel of Markov chain $\{\Psi_t^N\}_{t\in[\![0;N_0]\!]}$ is not *K-Lipschitz* in the sense of the definition (2.13) in [1]. However, the main arguments of the proof of Theorem 2 in [1] can still be applied to show that

$$\big\|V_t^N(\Psi_t^N)-\widehat{V}_t^{N,M}(\widehat{\Psi}_t^{N,M})\big\|_2\leq L_{V_t^N}\big\|\Psi_t^N-\widehat{\Psi}_t^{N,M}\big\|_2+\big\|V_{t+1}^N(\Psi_{t+1}^N)-\widehat{V}_{t+1}^{N,M}(\widehat{\Psi}_{t+1}^{N,M})\big\|_2,$$

for $t\in[\![0;N_0-1]\!]$ and

$$\big\|V_{N_0}^N(\Psi_0^N)-\widehat{V}_{N_0}^{N,M}(\widehat{\Psi}_{N_0}^{N,M})\big\|_2\leq L_{V_{N_0}^N}\big\|\Psi_{N_0}^N-\widehat{\Psi}_{N_0}^{N,M}\big\|_2,$$

where $L_{V_t^N}$ are given in Lemma 3.8. This implies that

$$\big|V_0^N(\Psi_0^N)-\widehat{V}_0^{N,M}(\widehat{\Psi}_0^{N,M})\big|\leq\sum_{t=0}^{N_0}L_{V_t^N}\|\Psi_t^N-\widehat{\Psi}_t^{N,M}\|_2.$$

Moreover, one has $V_0^N(\Psi_0^N)=\overline{\boldsymbol{\mathcal{H}}}_{\mathcal{M},N}(\mathbf{x},\mathbf{y})$ giving the result. □

Gathering together our three main results Theorems 2.2, 3.7, and 3.9, we obtain that the fully computable expression $\widehat{V}_0^{N,M}(\widehat{\Psi}_0^{N,M})$ is an approximation of our initial value function of interest $\overline{\boldsymbol{\mathcal{H}}}(\mathbf{x},\mathbf{y})$ with an error bound of

$$\big|\overline{\boldsymbol{\mathcal{H}}}(\mathbf{x},\mathbf{y})-\widehat{V}_0^{N,M}(\widehat{\Psi}_0^{N,M})\big|\leq K_1\epsilon_N\sum_{k=0}^{N_0-1}\|\mathfrak{B}^k\mathbf{H}\|_{\mathbb{L}(\mathcal{P}(\mathbb{X})\times\mathbb{Y})}+\sum_{t=0}^{N_0}L_{V_t^N}\|\Psi_t^N-\widehat{\Psi}_t^{N,M}\|_2,$$

that goes to zero as the number of points in the quantization grids goes to infinity.

# 4 Numerical example

In this section, we present a numerical example to illustrate our approximation results. It is adapted from the control of water tank problems which can be found in [3, section 1.3]. Such applications are essential in regions under high water stress.

Consider a water tank which capacity $K > 0$ is finite. It is filled with a random amount of rainfall each time it rains. However, the water level is only known through noisy measurements. One wants to cover the tank when the volume of water is closest to some value $\alpha \in (0; K)$. Let us model this situation with a $[0; K]^2$-valued finite-horizon Markov chain $(\tilde{\boldsymbol{\mathcal{X}}}_t, \tilde{\boldsymbol{\mathcal{Y}}}_t)_{t\in[\![0;N_0]\!]}$, where $(\tilde{\boldsymbol{\mathcal{X}}}_t)$ represents the sequence of water volumes contained in the tank and $(\tilde{\boldsymbol{\mathcal{Y}}}_t)$ symbolizes the measurements of $(\tilde{\boldsymbol{\mathcal{X}}}_t)$. We suppose that the dynamics of the Markov chain is given by

$$\left\{ \begin{array}{l} \tilde{\boldsymbol{\mathcal{X}}}_{t+1} = \min\left\{(\tilde{\boldsymbol{\mathcal{X}}}_t + \xi_t)_+; K\right\} \\ \tilde{\boldsymbol{\mathcal{Y}}}_{t+1} = \min\left\{(\tilde{\boldsymbol{\mathcal{X}}}_{t+1} + \psi_t)_+; K\right\} \end{array} \right.$$

where $x_+$ stands for the positive part of a real number $x$, and $(\xi_t)$ and $(\psi_t)$ are i.i.d. random variables with respective densities $f$ on $\mathbb{R}_+$ and $g$ on $\mathbb{R}$. Let us denote respectively $F$ and $G$ the cumulative distribution functions associated to $f$ and $g$. Let $B, C \in \mathcal{B}([0; K])$. The cost function is $\tilde{\mathbf{H}}(x, y) = K - |x - \alpha|$ so that the process is optimally stopped when the (unobserved) component $\tilde{\boldsymbol{\mathcal{X}}}_t$ is close or equal to $\alpha$. The transition law of this process is

$$\tilde{R}(B \times C|x, y) = \delta_0(B)F(-x)\mathfrak{M}_1(C) + \int_B f(\xi - x)\mathfrak{M}_2(C, \xi)d\xi + \delta_K(B)\mathfrak{M}_3(C)$$

where

$$\begin{aligned} \mathfrak{M}_1(C) &= \delta_0(C)G(0) + \int_C g(\psi)d\psi + \delta_K(C)(1 - G(K)), \\ \mathfrak{M}_2(C, \xi) &= \delta_0(C)G(-\xi) + \int_C g(\psi - \xi)d\psi + \delta_K(C)(1 - G(K - \xi)), \\ \mathfrak{M}_3(C) &= \delta_0(C)G(-K) + \int_C g(\psi - K)d\psi + \delta_K(C)(1 - G(0)). \end{aligned}$$

Assumption B does not hold when $[0; K]$ is endowed with the usual Euclidian norm because the points 0 and $K$ have a nonzero weight. Thus we change the topology to isolate these two points by adding an additional dimension to the process.

Consider the process $(\boldsymbol{\mathcal{X}}_t^1, \boldsymbol{\mathcal{X}}_t^2, \boldsymbol{\mathcal{Y}}_t)_{t\in[\![0;N_0]\!]}$, where $\boldsymbol{\mathcal{X}}_t^1 = \tilde{\boldsymbol{\mathcal{X}}}_t$, $\boldsymbol{\mathcal{Y}}_t = \tilde{\boldsymbol{\mathcal{Y}}}_t$ and the dynamics of $\boldsymbol{\mathcal{X}}_t^2$ is

$$\boldsymbol{\mathcal{X}}_{t+1}^2 = I_{\{\boldsymbol{\mathcal{X}}_{t+1}^1 = K\}} - I_{\{\boldsymbol{\mathcal{X}}_{t+1}^1 = 0\}}.$$

So, the unobservable state space is $\mathbb{X} = ((0; K) \times \{0\}) \cup \{(0, -1)\} \cup \{(K, 1)\}$. The observable state space is $\mathbb{Y} = [0; K]$. Let $\mathbf{H}(x_1, x_2, y) = K - |x_1 - \alpha|$ be the performance function. One may now write the transition law $R$ of the process $(\boldsymbol{\mathcal{X}}_t^1, \boldsymbol{\mathcal{X}}_t^2, \boldsymbol{\mathcal{Y}}_t)_{t\in[\![0;N_0]\!]}$ as

$$R(du_1, du_2, dv|x_1, x_2, y) = r(u_1, u_2, v, x_1, x_2, y)\lambda(du_1, du_2)\nu(dv),$$

where $r : (\mathbb{X} \times \mathbb{Y})^2 \to \mathbb{R}_+$ is defined by

$$\begin{aligned} & r(u_1, u_2, v, x_1, x_2, y) \\ &= 4I_{\{(0,-1)\}}(u_1, u_2)F(-x_1)\mathfrak{m}_1(v) + 2KI_{(0;K)\times\{0\}}(u_1, u_2)f(u_1 - x_1)\mathfrak{m}_2(u_1, v) \\ &\quad + 4I_{\{(K,1)\}}(u_1, u_2)(1 - F(K - x_1))\mathfrak{m}_3(v), \end{aligned}$$

with

$$\begin{aligned} \mathfrak{m}_1(v) &= 4G(0)I_{\{0\}}(v) + 2KI_{(0;K)}(v)g(v) + 4(1 - G(K))I_{\{K\}}(v), \\ \mathfrak{m}_2(u, v) &= 4G(-u)I_{\{0\}}(v) + 2KI_{(0;K)}(v)g(v - u) + 4(1 - G(K - u))I_{\{K\}}(v), \\ \mathfrak{m}_3(v) &= 4G(-K)I_{\{0\}}(v) + 2KI_{(0;K)}(v)g(v - K) + 4(1 - G(0))I_{\{K\}}(v), \end{aligned}$$

and

$$\begin{aligned}\lambda(du_1, du_2) &= \frac{\delta_0(du_1)\delta_{-1}(du_2)}{4} + \frac{\mu(du_1)\delta_0(du_2)}{2K} + \frac{\delta_K(du_1)\delta_1(du_2)}{4},\\ \nu(dv) &= \frac{\delta_0(dv)}{4} + \frac{\mu(dv)}{2K} + \frac{\delta_K(dv)}{4},\end{aligned}$$

where $\mu$ denotes the Lebesgue measure. One may note that neither $\lambda$ nor $\nu$ are absolutely continuous with respect to the Lebesgue measure on $[0; K]$. Thus, this model does not satisfy the assumptions of [8, 9, 10]. However, our Assumption A is clearly satisfied.

Assume that $f$ is Lipschitz-continuous on $[0; K]$ with constant $L_f$ and $g$ is positive and Lipschitz-continuous on $[-K; K]$ with constant $L_g$ (e.g. if $f$ is an exponential density function and $g$ a centered Gaussian density function, these hold). Therefore, they are both bounded above on these intervals, respectively by $\|f\|_{sup}$ and $\|g\|_{sup}$. Straightforward calculations show that assumptions B and D hold with the following constants

$$\begin{aligned}\overline{r} &= (8 + 2K\|f\|_{sup})(8 + 2K\|g\|_{sup})\\ L_r &= \max\Big\{\mathfrak{a}, \mathfrak{b}, \frac{8(8 + 2K\|g\|_{sup})}{K + 2}, (8 + 2K\|g\|_{sup})(8\|f\|_{sup} + 2KL_f)\Big\},\\ \|\mathbf{H}\|_{\mathbb{L}(\mathbb{X}\times\mathbb{Y})} &\leq K + 1,\end{aligned}$$

where

$$\mathfrak{a} = 2K(L_f(8 + 2K\|g\|_{sup}) + \|f\|_{sup}(8\|g\|_{sup} + 2KL_g))$$

and

$$\mathfrak{b} = \max\big\{2K\|f\|_{sup}(8\|g\|_{sup} + 2KL_g); (4 + 2K\|f\|_{sup} + 2KL_f)(8 + 2K\|g\|_{sup})\big\}.$$

Assumption C requires that the density $g$ be bounded from below by some positive number $\mathfrak{g}$ on $[-K; K]$. Thus, one has

$$\mathfrak{m}_3(v) \geq \min\{4G(-K); 2K\mathfrak{g}; 4(1 - G(0))\}.$$

As

$$r(\lambda, v, x_1, x_2, y) = \int_{\mathbb{X}} r(u_1, u_2, v, x_1, x_2, y)\lambda(du_1, du_2) \geq (1 - F(K - x_1))\mathfrak{m}_3(v),$$

let us suppose that $F(K) < 1$, $G(0) < 1$ and $G(-K) > 0$. These are verified by exponential and centered Gaussian density functions as above, for instance. One then deduces that

$$r(\lambda, v, x_1, x_2, y) \geq (1 - F(K)) \min\big(4G(-K); 2K\mathfrak{g}; 4(1 - G(0))\big) > 0.$$

for all $v, y \in \mathbb{Y}$ and $(x_1, x_2) \in \mathbb{X}$. Therefore, this shows that assumption C holds.

For our numerical experimentations, we chose $N_0 = 10$, $K = 1$, $\alpha = 0.5$, and the initial state $(\boldsymbol{\mathcal{X}}_0^1, \boldsymbol{\mathcal{X}}_0^2, \boldsymbol{\mathcal{Y}}_0) = (0, -1, 0)$. We suppose that the $\xi_t$ are exponentially distributed with parameter 5. We suppose that the $\psi_t$ are normally distributed with mean 0 and standard deviation 0.03. Following the method developed in this paper, we have performed the two quantizations by using the competitive learning vector quantization algorithm (see section 2.2 of [6]). Table 1 displays the approximation $\widehat{V}_0^{N,M}(\widehat{\Psi}_0^{N,M})$ of the value function at $(0, -1, 0)$ according to the numbers $N$ and $M$ of points in the quantization grids. The exact value function is not known, but as expected one sees that our approximation is close to the optimal performance of 1.

| $M$ | $N=12$ | $N=25$ | $N=50$ | $N=100$ |
|---|---|---|---|---|
| 125 | 0.9323 | 0.9534 | | |
| 250 | 0.9381 | 0.9579 | | |
| 500 | 0.9392 | 0.9577 | | |
| 1000 | 0.9404 | 0.9574 | | |
| 10000 | 0.9416 | 0.9578 | 0.9686 | 0.9771 |

Table 1: Approximation $\widehat{V}_0^{N,M}(\widehat{\Psi}_0^{N,M})$ of the optimal value according to $M$ and $N$

# Appendix A: Proof of Theorem 2.2

In order to prove Theorem 2.2, we need to introduce a new auxiliary control model $\boldsymbol{\mathcal{M}}$ given by the five-tuple $\big(\mathbb{F}, \mathbb{A}, T, H, h\big)$ where

(a) the state space is $\mathbb{F} = \mathbb{X} \times \mathbb{Y} \times \{0,1\}$,

(b) the action space is $\mathbb{A} = \{0,1\}$,

(c) the transition probability function is given the stochastic kernel $T$ on $\mathbb{F}$ given $\mathbb{F} \times \mathbb{A}$ defined by $T(B \times C|x,y,z,a) = R(B \times C|x,y)\big[\delta_z(D)I_{\{a=0\}} + \delta_1(D)I_{\{a=1\}}\big]$ for any any $B \in \mathcal{B}(\mathbb{X})$, $C \in \mathcal{B}(\mathbb{Y})$, $D \subset \{0,1\}$ and $(x,y,z,a) \in \mathbb{F} \times \mathbb{A}$,

(d) the cost-per-stage $H$ and the terminal cost $h$.

Define $\boldsymbol{\Omega} = \mathbb{F}^{N_0+1}$ and $\boldsymbol{\mathcal{F}}$ its associated product $\sigma$-algebra. Introduce the coordinate projections $\mathbf{X}_t$ (respectively $\mathbf{Y}_t$, and $\mathbf{Z}_t$) from $\boldsymbol{\Omega}$ to the set $\mathbb{X}$ (respectively $\mathbb{Y}$, and $\{0,1\}$). Consider an arbitrary policy $\pi \in \Pi^o$. Define recursively the action process $\{\mathbf{A}_t\}_{t\in[\![0;N_0-1]\!]}$ by $\mathbf{A}_t = \pi_t(\mathbf{Y}_0, \mathbf{Z}_0, \mathbf{A}_0, \ldots, \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \mathbf{A}_{t-1}, \mathbf{Y}_t, \mathbf{Z}_t)$ for $t \in [\![1; N_0-1]\!]$ and $\mathbf{A}_0 = \pi_0(\mathbf{Y}_0, \mathbf{Z}_0)$. Define the filtration $\{\boldsymbol{\mathcal{F}}_t\}_{t\in[\![0;N_0]\!]}$ by $\boldsymbol{\mathcal{F}}_t = \sigma\{\mathbf{X}_0, \mathbf{Y}_0, \mathbf{Z}_0, \ldots, \mathbf{X}_t, \mathbf{Y}_t, \mathbf{Z}_t\}$ for $t \in [\![0; N_0]\!]$. According to [2, 4], there exists a probability measure $\mathbf{P}^\pi_{(\mathbf{x},\mathbf{y})}$ on $\big(\boldsymbol{\Omega}, \boldsymbol{\mathcal{F}}\big)$ satisfying

i) $\mathbf{P}^\pi_{(\mathbf{x},\mathbf{y})}\big((\mathbf{X}_0, \mathbf{Y}_0, \mathbf{Z}_0) \in B \times C \times D\big) = \delta_{(\mathbf{x},\mathbf{y})}(B \times C)\delta_0(D)$,

ii) $\mathbf{P}^\pi_{(\mathbf{x},\mathbf{y})}\big((\mathbf{X}_{t+1}, \mathbf{Y}_{t+1}, \mathbf{Z}_{t+1}) \in B \times C \times D|\boldsymbol{\mathcal{F}}_t\big) = T(B \times C \times D|\mathbf{X}_t, \mathbf{Y}_t, \mathbf{Z}_t, \mathbf{A}_t)$,

for $t \in [\![0; N_0-1]\!]$, $B \in \mathcal{B}(\mathbb{X})$, $C \in \mathcal{B}(\mathbb{Y})$, $D \subset \{0,1\}$.

The expectation under the probability $\mathbf{P}^\pi_{(\mathbf{x},\mathbf{y})}$ is denoted by $\mathbf{E}^\pi_{(\mathbf{x},\mathbf{y})}$. For a policy $\pi \in \Pi^o$, the performance criterion is given by

$$\boldsymbol{\mathcal{H}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x},\mathbf{y},\pi) = \mathbf{E}^\pi_{(\mathbf{x},\mathbf{y})}\Big[\sum_{t=0}^{N_0-1} H(\mathbf{X}_t, \mathbf{Y}_t, \mathbf{Z}_t, \mathbf{A}_t)\Big] + \mathbf{E}^\pi_{(\mathbf{x},\mathbf{y})}\big[h(\mathbf{X}_{N_0}, \mathbf{Y}_{N_0}, \mathbf{Z}_{N_0})\big]. \tag{35}$$

The optimization problem we are interested in is to maximize the reward function $\boldsymbol{\mathcal{H}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x},\mathbf{y},\pi)$ over $\Pi^o$ and $\overline{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x},\mathbf{y}) = \sup_{\pi\in\Pi^o} \boldsymbol{\mathcal{H}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x},\mathbf{y},\pi)$. We first need to prove the following technical lemma.

**Lemma A.1** *For any $t \in [\![0; N_0]\!]$,*

$$\sigma\{\mathbf{Y}_0, \mathbf{Z}_0, \ldots, \mathbf{Y}_t, \mathbf{Z}_t\} = \sigma\{\mathbf{Y}_0, \ldots, \mathbf{Y}_t\}.$$

**Proof:** Clearly, $\mathbf{A}_t$ is measurable with respect to $\sigma\{\mathbf{Y}_0, \mathbf{Z}_0, \ldots, \mathbf{Y}_t, \mathbf{Z}_t\}$ for $t \in [\![0; N_0 - 1]\!]$. Moreover, from the definition of the transition kernel $T$, we obtain that $\mathbf{Z}_t = I_{\{\mathbf{A}_{t-1}=1\}} + \mathbf{Z}_{t-1} I_{\{\mathbf{A}_{t-1}=0\}}$ for any $t \in [\![1; N_0]\!]$. Recalling that $\mathbf{Z}_0 = 0$, it follows easily $\sigma\{\mathbf{Y}_0, \mathbf{Z}_0, \ldots, \mathbf{Y}_t, \mathbf{Z}_t\} \subset \sigma\{\mathbf{Y}_0, \ldots, \mathbf{Y}_t\}$ for $t \in [\![0; N_0]\!]$ showing the result. □

The next result shows that the optimization problem defined through $\boldsymbol{\mathcal{M}}$ is equivalent to the initial optimal stopping problem defined in Definition 2.1.

**Proposition A.2** *The following assertions hold.*
*i) For any control $\ell \in L$, there exist a policy $\pi \in \Pi^o$ such that*

$$\boldsymbol{\mathcal{H}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x}, \mathbf{y}, \pi) = \boldsymbol{\mathcal{H}}(\mathbf{x}, \mathbf{y}, \ell).$$

*ii) For any policy $\pi \in \Pi^o$, there exist a control $\ell \in L$ such that*

$$\boldsymbol{\mathcal{H}}(\mathbf{x}, \mathbf{y}, \ell) = \boldsymbol{\mathcal{H}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x}, \mathbf{y}, \pi).$$

**Proof:** Regarding item *i)*, consider a control $\ell = \big(\boldsymbol{\Xi}, \boldsymbol{\mathcal{G}}, \mathbf{Q}, \{\boldsymbol{\mathcal{G}}_t\}_{t\in[\![0;N_0]\!]}, \{\boldsymbol{\mathcal{X}}_t, \boldsymbol{\mathcal{Y}}_t\}_{t\in[\![0;N_0]\!]}, \tau\big)$ in $L$. On the probability space $\big(\boldsymbol{\Xi}, \boldsymbol{\mathcal{G}}, \mathbf{Q}\big)$, let us define the processes $\{\boldsymbol{\mathcal{A}}_t\}_{t\in[\![0;N_0-1]\!]}$ and $\{\boldsymbol{\mathcal{Z}}_t\}_{t\in[\![0;N_0]\!]}$ by $\boldsymbol{\mathcal{A}}_t = I_{\{\tau\le t\}}$ and $\boldsymbol{\mathcal{Z}}_t = \boldsymbol{\mathcal{A}}_{t-1}$ for $t \in [\![1; N_0]\!]$ and $\boldsymbol{\mathcal{Z}}_0 = 0$. Introduce the filtrations $\{\boldsymbol{\mathcal{T}}_t\}_{t\in[\![0;N_0]\!]}$ by $\boldsymbol{\mathcal{T}}_t = \sigma\{\boldsymbol{\mathcal{X}}_0, \boldsymbol{\mathcal{Y}}_0, \boldsymbol{\mathcal{Z}}_0, \boldsymbol{\mathcal{A}}_0, \ldots, \boldsymbol{\mathcal{X}}_t, \boldsymbol{\mathcal{Y}}_t, \boldsymbol{\mathcal{Z}}_t, \boldsymbol{\mathcal{A}}_t\}$ and $\{\boldsymbol{\mathcal{G}}_t^{\boldsymbol{\mathcal{Y}}}\}_{t\in[\![0;N_0]\!]}$ by $\boldsymbol{\mathcal{G}}_t^{\boldsymbol{\mathcal{Y}}} = \sigma\{\boldsymbol{\mathcal{Y}}_0, \ldots, \boldsymbol{\mathcal{Y}}_t\}$. Since $\tau$ is an $\{\boldsymbol{\mathcal{G}}_t^{\boldsymbol{\mathcal{Y}}}\}_{t\in[\![0;N_0]\!]}$-stopping time, we have $\boldsymbol{\mathcal{T}}_t \subset \boldsymbol{\mathcal{G}}_t$. Moreover, $\boldsymbol{\mathcal{Z}}_{t+1}$ is $\boldsymbol{\mathcal{T}}_t$-measurable. Consequently, it is easy to show that

$$\mathbf{Q}\big((\boldsymbol{\mathcal{X}}_{t+1}, \boldsymbol{\mathcal{Y}}_{t+1}, \boldsymbol{\mathcal{Z}}_{t+1}) \in B \times C \times D | \boldsymbol{\mathcal{T}}_t\big) = I_{\{\boldsymbol{\mathcal{Z}}_{t+1}\in D\}} R(B \times C | \boldsymbol{\mathcal{X}}_t, \boldsymbol{\mathcal{Y}}_t).$$

We have $\{\boldsymbol{\mathcal{A}}_t = 1\} = \{\boldsymbol{\mathcal{Z}}_{t+1} = 1\}$ and $\{\boldsymbol{\mathcal{A}}_t = 0\} = \{\boldsymbol{\mathcal{Z}}_{t+1} = 0\} \subset \{\boldsymbol{\mathcal{A}}_{t-1} = 0\} = \{\boldsymbol{\mathcal{Z}}_t = 0\}$, and so

$$\begin{aligned}\mathbf{Q}\big((\boldsymbol{\mathcal{X}}_{t+1}, \boldsymbol{\mathcal{Y}}_{t+1}, \boldsymbol{\mathcal{Z}}_{t+1}) \in B \times C \times D | \boldsymbol{\mathcal{T}}_t\big) =& \big[I_{\{\boldsymbol{\mathcal{A}}_t=0\}} \delta_{\{\boldsymbol{\mathcal{Z}}_t\in D\}} + I_{\{\boldsymbol{\mathcal{A}}_t=1\}} \delta_1(D)\big] R(B \times C | \boldsymbol{\mathcal{X}}_t, \boldsymbol{\mathcal{Y}}_t) \\ =& T(B \times C \times D | \boldsymbol{\mathcal{X}}_t, \boldsymbol{\mathcal{Y}}_t, \boldsymbol{\mathcal{Z}}_t, \boldsymbol{\mathcal{A}}_t). \end{aligned} \tag{36}$$

Now, there exists an $\mathbb{A}$-valued measurable mapping $\pi_t$ defined on $\mathbb{Y}^{t+1}$ satisfying $\boldsymbol{\mathcal{A}}_t = \pi_t(\boldsymbol{\mathcal{Y}}_0, \ldots, \boldsymbol{\mathcal{Y}}_t)$ and so,

$$\mathbf{Q}(\boldsymbol{\mathcal{A}}_t \in F | \sigma\{\boldsymbol{\mathcal{Y}}_0, \boldsymbol{\mathcal{Z}}_0, \boldsymbol{\mathcal{A}}_0, \ldots, \boldsymbol{\mathcal{Y}}_t, \boldsymbol{\mathcal{Z}}_t\}) = \delta_{\pi_t(\boldsymbol{\mathcal{Y}}_0, \ldots, \boldsymbol{\mathcal{Y}}_t)}(F), \tag{37}$$

for any $t \in [\![0; N_0 - 1]\!]$ and $F \subset \mathbb{A}$. Recall that

$$\mathbf{Q}\big((\boldsymbol{\mathcal{X}}_0, \boldsymbol{\mathcal{Y}}_0, \boldsymbol{\mathcal{Z}}_0) \in B \times C \times D\big) = \delta_{(\mathbf{x},\mathbf{y})}(B \times C)\delta_0(D) \tag{38}$$

for any $B \in \mathcal{B}(\mathbb{X})$, $C \in \mathcal{B}(\mathbb{Y})$, $D \subset \{0, 1\}$. Combining equations (36)-(38) and by the uniqueness property in the Theorem of Ionescu-Tulcea (see, e.g. [4, Proposition C.10]), it follows that for the control policy $\pi = \{\pi_t\}_{t\in[\![0;N_0]\!]}$

$$\begin{aligned}&\mathbf{Q}\big((\boldsymbol{\mathcal{X}}_0, \boldsymbol{\mathcal{Y}}_0, \boldsymbol{\mathcal{Z}}_0, \boldsymbol{\mathcal{A}}_0, \ldots, \boldsymbol{\mathcal{X}}_{N_0-1}, \boldsymbol{\mathcal{Y}}_{N_0-1}, \boldsymbol{\mathcal{Z}}_{N_0-1}, \boldsymbol{\mathcal{A}}_{N_0-1}, \boldsymbol{\mathcal{X}}_{N_0}, \boldsymbol{\mathcal{Y}}_{N_0}, \boldsymbol{\mathcal{Z}}_{N_0}) \in H\big) \\ &\quad = \mathbf{P}^{\pi}_{(\mathbf{x},\mathbf{y})}\big((\mathbf{X}_0, \mathbf{Y}_0, \mathbf{Z}_0, \mathbf{A}_0, \ldots, \mathbf{X}_{N_0-1}, \mathbf{Y}_{N_0-1}, \mathbf{Z}_{N_0-1}, \mathbf{A}_{N_0-1}, \mathbf{X}_{N_0}, \mathbf{Y}_{N_0}, \mathbf{Z}_{N_0}) \in H\big) \end{aligned} \tag{39}$$

for any $H \in \boldsymbol{\mathcal{F}}$.

Observe that for $k \in [\![0; N_0 - 1]\!]$ we have $\{\tau = k\} = \{\mathbf{Z}_k = 0\} \cup \{\mathbf{A}_k = 1\}$ and $\{\tau = N_0\} = \{\mathbf{Z}_{N_0} = 0\}$. Consequently,

$$\begin{aligned}\mathbf{E}^{\mathbf{Q}}_{(\mathbf{x},\mathbf{y})}\big[\mathbf{H}(\boldsymbol{\mathcal{X}}_\tau, \boldsymbol{\mathcal{Y}}_\tau)\big] &= \sum_{t=0}^{N_0-1} \mathbf{E}^{\mathbf{Q}}_{(\mathbf{x},\mathbf{y})}\big[\mathbf{H}(\boldsymbol{\mathcal{X}}_t, \boldsymbol{\mathcal{Y}}_t) I_{\{\tau=t\}}\big] + \mathbf{E}^{\mathbf{Q}}_{(\mathbf{x},\mathbf{y})}\big[\mathbf{H}(\boldsymbol{\mathcal{X}}_{N_0}, \boldsymbol{\mathcal{Y}}_{N_0}) I_{\{\tau=N_0\}}\big] \\ &= \sum_{t=0}^{N_0-1} \mathbf{E}^{\mathbf{Q}}_{(\mathbf{x},\mathbf{y})}\big[\mathbf{H}(\boldsymbol{\mathcal{X}}_t, \boldsymbol{\mathcal{Y}}_t) I_{\{(\boldsymbol{\mathcal{Z}}_t, \boldsymbol{\mathcal{A}}_t)=(0,1)\}}\big] + \mathbf{E}^{\mathbf{Q}}_{(\mathbf{x},\mathbf{y})}\big[\mathbf{H}(\boldsymbol{\mathcal{X}}_{N_0}, \boldsymbol{\mathcal{Y}}_{N_0}) I_{\{\boldsymbol{\mathcal{Z}}_{N_0}=1\}}\big]\end{aligned}$$

Now, by using the definitions of $H$ and $h$ we get

$$\mathbf{E}^{\mathbf{Q}}_{(\mathbf{x},\mathbf{y})}\big[\mathbf{H}(\boldsymbol{\mathcal{X}}_\tau,\boldsymbol{\mathcal{Y}}_\tau)\big] = \sum_{t=0}^{N_0-1}\mathbf{E}^{\mathbf{Q}}_{(\mathbf{x},\mathbf{y})}\big[H(\boldsymbol{\mathcal{X}}_t,\boldsymbol{\mathcal{Y}}_t,\boldsymbol{\mathcal{Z}}_t,\boldsymbol{\mathcal{A}}_t)\big] + \mathbf{E}^{\mathbf{Q}}_{(\mathbf{x},\mathbf{y})}\big[h(\boldsymbol{\mathcal{X}}_{N_0},\boldsymbol{\mathcal{Y}}_{N_0},\boldsymbol{\mathcal{Z}}_{N_0})\big].$$

By using equation (39), it follows that

$$\mathbf{E}^{\mathbf{Q}}_{(\mathbf{x},\mathbf{y})}\big[\mathbf{H}(\boldsymbol{\mathcal{X}}_\tau,\boldsymbol{\mathcal{Y}}_\tau)\big] = \sum_{t=0}^{N_0-1}\mathbf{E}^{\pi}_{(\mathbf{x},\mathbf{y})}\big[H(\mathbf{X}_t,\mathbf{Y}_t,\mathbf{Z}_t,\mathbf{A}_t)\big] + \mathbf{E}^{\pi}_{(\mathbf{x},\mathbf{y})}\big[h(\mathbf{X}_{N_0},\mathbf{Y}_{N_0},\mathbf{Z}_{N_0})\big],$$

showing the first claim.

Regarding item *ii*), let $\pi$ be a policy in $\Pi^o$. Then, on the probability space $\big(\boldsymbol{\Omega},\boldsymbol{\mathcal{F}},\mathbf{P}^{\pi}_{(\mathbf{x},\mathbf{y})}\big)$, $\{\mathbf{X}_t,\mathbf{Y}_t\}_{t\in[\![0;N_0]\!]}$ is an $\{\boldsymbol{\mathcal{F}}_t\}_{t\in[\![0;N_0]\!]}$-adapted Markov chain with transition kernel $R$ and with initial distribution $\delta_{(\mathbf{x},\mathbf{y})}$. Introduce the $[\![0;N_0]\!]$-valued random variable $\tau$ defined by

$$\tau = \begin{cases}\inf\{k\in[\![0;N_0-1]\!]:\mathbf{A}_k=1\} & \text{if } \{k\in[\![0;N_0-1]\!]:\mathbf{X}_k=1\}\neq\emptyset,\\ N_0 & \text{otherwise.}\end{cases}$$

It follows from Lemma A.1 that $\tau$ is a stopping time with respect to $\big\{\sigma\{\mathbf{Y}_0,\ldots,\mathbf{Y}_t\}_{t\in[\![0;N_0]\!]}\big\}$ showing that the control $\lambda$ defined by $\Big(\boldsymbol{\Omega},\boldsymbol{\mathcal{F}},\mathbf{P}^{\pi}_{(\mathbf{x},\mathbf{y})},\{\boldsymbol{\mathcal{F}}_t\}_{t\in[\![0;N_0]\!]},\{\mathbf{X}_t,\mathbf{Y}_t\}_{t\in[\![0;N_0]\!]},\tau\Big)$ belongs to $\Lambda$. Recalling that $\mathbf{Z}_0=0$ and that $\mathbf{Z}_t = I_{\{\mathbf{A}_{t-1}=1\}}+\mathbf{Z}_{t-1}I_{\{\mathbf{A}_{t-1}=0\}}$ for any $t\in[\![1;N_0]\!]$, we get that $\{\tau=t\}=\{\mathbf{Z}_t=0\}\cup\{\mathbf{A}_t=1\}$ for $t\in[\![0;N_0-1]\!]$ and $\{\tau=N_0\}=\{\mathbf{Z}_{N_0}=0\}$. Now, by using the definitions of $H$ and $h$ it follows that

$$\begin{aligned}&\sum_{t=0}^{N_0-1}\mathbf{E}^{\pi}_{(\mathbf{x},\mathbf{y})}\big[H(\mathbf{X}_t,\mathbf{Y}_t,\mathbf{Z}_t,\mathbf{A}_t)\big] + \mathbf{E}^{\pi}_{(\mathbf{x},\mathbf{y})}\big[h(\mathbf{X}_{N_0},\mathbf{Y}_{N_0},\mathbf{Z}_{N_0})\big]\\ &\qquad= \sum_{t=0}^{N_0-1}\mathbf{E}^{\pi}_{(\mathbf{x},\mathbf{y})}\big[\mathbf{H}(\mathbf{X}_t,\mathbf{Y}_t)I_{\{(\mathbf{Z}_t,\mathbf{A}_t)=(0,1)\}}\big] + \mathbf{E}^{\pi}_{(\mathbf{x},\mathbf{y})}\big[\mathbf{H}(\mathbf{X}_{N_0},\mathbf{Y}_{N_0})I_{\{\mathbf{Z}_{N_0}=1\}}\big]\\ &\qquad= \mathbf{E}^{\pi}_{(\mathbf{x},\mathbf{y})}\big[\mathbf{H}(\mathbf{X}_\tau,\mathbf{Y}_\tau)\big],\end{aligned}$$

implying that $\boldsymbol{\mathcal{H}}(\mathbf{x},\mathbf{y},\ell)=\boldsymbol{\mathcal{H}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x},\mathbf{y},\pi)$ and showing the second claim. □

**Proof of Theorem 2.2** From Theorem 5.3.2 in [2] we get that $\overline{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\mathcal{M}}}(\mathbf{x},\mathbf{y})=\overline{\boldsymbol{\mathcal{H}}}_{\mathcal{M}}(\mathbf{x},\mathbf{y})$ and so from Proposition A.2, it follows that $\overline{\boldsymbol{\mathcal{H}}}(\mathbf{x},\mathbf{y})=\overline{\boldsymbol{\mathcal{H}}}_{\mathcal{M}}(\mathbf{x},\mathbf{y})$ giving the first equality in equation (7). Under Assumptions (A1), B and D, the hypotheses of Theorems 5.3.3 in [2] are satisfied. Therefore, it follows that the Bellman equation $\{v_k\}_{k\in[\![0;N_0]\!]}$ for the model $\mathcal{M}$ is given by

$$\begin{cases}v_0(\theta,y,z)=h(\theta,y,z)\\ v_k(\theta,y,z)=\max_{a\in\mathbb{A}}\big\{H(\theta,y,z,a)+Qv_{k-1}(\theta,y,z,a)\big\}\end{cases}$$

and satisfies $\overline{\boldsymbol{\mathcal{H}}}_{\mathcal{M}}(\mathbf{x},\mathbf{y})=v_{N_0}(\delta_{\mathbf{x}},\mathbf{y},0)$. However, since $h(\theta,y,1)=H(\theta,y,1)=0$, it easy to show that $v_k(\theta,y,1)=0$ for any $(\theta,y)\in\mathcal{P}(\mathbb{X})\times\mathbb{Y}$ and $k\in[\![0;N_0]\!]$. Moreover, by using the definitions of $h$, $H$ and the kernel $Q$ we obtain that $v_0(\theta,y,0)=\mathbf{H}(\theta,y)$ and

$$\begin{aligned}v_k(\theta,y,0)&=\max_{a\in\mathbb{A}}\big\{H(\theta,y,0,a)+Qv_{k-1}(\theta,y,0,a)\big\}\\ &=\max\big\{\mathbf{H}(\theta,y),Sv_{k-1}(\theta,y)\big\}=\mathfrak{B}v_{k-1}(\theta,y)\end{aligned}$$

for any $(\theta,y)\in\mathcal{P}(\mathbb{X})\times\mathbb{Y}$ and $k\in[\![1;N_0]\!]$ implying that $\overline{\boldsymbol{\mathcal{H}}}_{\mathcal{M}}(\mathbf{x},\mathbf{y})=\mathfrak{B}^{N_0}\mathbf{H}(\delta_{\mathbf{x}},\mathbf{y})$ and giving the second equality in equation (7). □

## References


[1] Vlad Bally, Gilles Pagès, and Jacques Printems. A quantization tree method for pricing and hedging multidimensional American options. *Math. Finance*, 15(1):119–168, 2005.

[2] Nicole Bäuerle and Ulrich Rieder. *Markov decision processes with applications to finance*. Universitext. Springer, Heidelberg, 2011.

[3] Onésimo Hernández-Lerma. *Adaptive Markov control processes*, volume 79 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1989.

[4] Onésimo Hernández-Lerma and Jean Bernard Lasserre. *Discrete-time Markov control processes*, volume 30 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1996.

[5] Gilles Pagès, Huyên Pham, and Jacques Printems. An optimal Markovian quantization algorithm for multi-dimensional stochastic control problems. *Stoch. Dyn.*, 4(4):501–545, 2004.

[6] Gilles Pagès, Huyên Pham, and Jacques Printems. Optimal quantization methods and applications to numerical problems in finance. In *Handbook of computational and numerical methods in finance*, pages 253–297. Birkhäuser Boston, Boston, MA, 2004.

[7] Huyên Pham, Wolfgang Runggaldier, and Afef Sellami. Approximation by quantization of the filter process and applications to optimal stopping problems under partial observation. *Monte Carlo Methods Appl.*, 11(1):57–81, 2005.

[8] Fan Ye and Enlu Zhou. Optimal stopping of partially observable markov processes: A filtering-based duality approach. *Automatic Control, IEEE Transactions on*, 58(10):2698–2704, 2013.

[9] Enlu Zhou. Optimal stopping under partial observation: Near-value iteration. *Automatic Control, IEEE Transactions on*, 58(2):500–506, 2013.

[10] Enlu Zhou, M.C. Fu, and S.I. Marcus. Solving continuous-state pomdps via density projection. *Automatic Control, IEEE Transactions on*, 55(5):1101–1116, 2010.